The Bell System Technical Journal

Devoted to the Scientific and Engineering Aspects of Electrical Communication

April, 1923

Impedance of Smooth Lines, and Design of Simulating Networks

By RAY S. HOYT

Introduction

THE transmission of alternating currents over any transmission line between specified terminal impedances depends only on the propagation constant and the characteristic impedance of the line (at the particular frequency contemplated). In this sense, then, the properties of transmission lines may be classed broadly as propagation characteristics and impedance characteristics. In telephony we are primarily concerned with the dependence of these characteristics on the frequency, over the telephonic frequency range.

Prior to the application of telephone repeaters to telephone lines the propagation characteristics of such lines were more important than their impedance characteristics, because the received energy depended much more on the former than on the latter.

The application of the two-way telephone repeater greatly altered the relative importance of these two characteristics, decreasing the need for high transmitting efficiency of a line but greatly increasing the dependence of the results on the impedance of the line. As well known, this is because the amplification to which a two-way repeater can be set without singing, or even without serious injury to the intelligibility of the transmission, depends strictly on the degree of impedance-balance between the lines or between the lines and their balancing-networks. In the case of the 21-type repeater the two lines must have such impedances as to closely balance each other throughout the telephonic frequency range. In the case of the 22-type repeater, which for long lines requiring more than one repeater is superior to the

21-type, impedance networks are required for closely balancing the impedances of the two lines throughout the telephonic frequency range. Such balancing networks are necessary also in connection with the so-called four-wire repeater circuit.[1]

Smooth lines are fundamental in telephonic transmission; for any telephone line is either a simple smooth line, or a compound smooth line, or a periodically loaded line whose sections are themselves short smooth lines. In any case the characteristic impedances of the constituent smooth lines enter importantly into the impedance of the system. Moreover, the characteristic impedance of the series type of periodically loaded line, at frequencies low relatively to its critical frequency, is closely the same as the characteristic impedance of the corresponding smooth line.

Parts I, II, and III of this paper aim to present in a simple yet comprehensive manner the dependence of the characteristic impedance of the various types of smooth lines on the frequency and on the line constants, by means of description accompanied by equations transformed to the most suitable forms and by graphs of such equations.

Part IV describes the principal networks devised by the writer at various times within about the last ten years, for simulating the impedance of the various types of smooth lines. Of course, the impedance of any line could be simulated, as closely as desired, by means of an artificial model constructed of many short sections each having lumped constants; but such structures would be very expensive and very cumbersome. Compared with them the networks described in this paper are very simple non-periodic structures that are relatively inexpensive and are quite compact; yet the more precise of them have proved to be adequate for simulating with high precision the impedance of most types of smooth lines, while even the least precise (which are the simplest) suffice for a good many applications. The paper includes first approximation design-formulas and outlines a supplementary semi-graphical method for arriving at the best proportioning of the networks. A typical illustrative example is worked out in Appendix E.

It is hoped to devote a succeeding paper to the impedance characteristics of periodically loaded lines, and to various networks devised for simulating and compensating such impedance.

[1]Regarding the broad subject of repeaters and repeater circuits, reference may be made to the paper by Gherardi and Jewett: "Telephone Repeaters," A. I. E. E. Trans., 1919, pp. 1287–1345.

PART I

GENERAL CONSIDERATIONS PERTAINING TO SMOOTH LINES

The exact formula for the characteristic impedance K of any smooth line is usually written in the form

$$K=\sqrt{\frac{R+i\omega L}{G+i\omega C}} \tag{1}$$

R, G, L and C denoting, as usual, the fundamental line constants,[2] namely, the resistance, leakance (leakage conductance), inductance, and capacity, per unit length; ω denoting 2π times the frequency f; and i the imaginary operator $\sqrt{-1}$.

However, this form is neither the simplest nor the most significant. For it involves separately the four quantities R, G, L, and C and is thus a function of not less than four[3] variables, whereas its value evidently depends on only the relative values of these quantities and hence must be expressible as a function of only three independent variables—namely the ratios of any three of them to the fourth.

In deciding just what form or forms of expression to adopt for K we shall here be guided by the following practical considerations:

(A) In telephony we are chiefly interested in the dependence of K on the frequency f; or, stated more generally, in the dependence of some quantity that is approximately proportional to K on some quantity that is approximately proportional to f.

The class of smooth lines is comprised between the following two rather wide extremes, having very different characteristics:

(B) At one extreme are the large gauge open-wire lines, particularly when used at high frequencies. For them R is small relatively to ωL, and G relatively to ωC; and hence K is approximately or at least roughly equal to $\sqrt{L/C}$.

(C) At the other extreme are the small gauge cables, particularly when used at low frequencies. For them R is large relatively to ωL, though G is small relatively to ωC; and hence K is approximately or at least roughly equal to $\sqrt{R/i\omega C}$.

(D) The line constants R, L, C do not change much with frequency over at least the voice frequency range; and hence they, or combinations of them, serve suitably as parameters.

(E) The leakance G, which is nearly always the least important of the four line constants, usually varies greatly with the frequency

[2] Constants as to current and voltage.

[3] Five if ω is regarded separately from L and C.

and hence by itself does not serve very suitably as a parameter. However, in a wide range of applications G is approximately or at least roughly proportional to the frequency; and then a suitable parameter is G/f or preferably $G/\omega C$. This is true of cables except at extremely low frequencies. It is at least roughly true of open-wire lines at very high frequencies, such as carrier frequencies, but usually not at voice frequencies. For most lines the leakance G is usually approximately or at least roughly a linear function of the frequency, namely, $G=G_0+\nu f$, where G_0 is the leakance at $f=0$, and ν is approximately independent of the frequency. For cables, G_0 is small compared with νf except at very low values of f; but for open-wire lines G_0 is usually not negligible except at high values of f.

In the light of these considerations a study of equation (1) suggests the employment of the quantities F, E, k, g, a, b defined by the following six equations. Not all of these substitutions will be employed simultaneously, but it is convenient to set them all down here together.

$$F=\omega L/R, \qquad (2) \qquad\qquad E=\omega C/R, \qquad (3)$$
$$k=\sqrt{L/C}, \qquad (4) \qquad\qquad g=\sqrt{G/R}, \qquad (5)$$
$$a=GL/RC, \qquad (6) \qquad\qquad b=G/\omega C. \qquad (7)$$

Usually F or E will be treated as the independent variable; and k, g, a, b as parameters.

It should perhaps here be emphasized that the approximations mentioned in the foregoing set of five considerations, (A) to (E), are employed merely as guides in the selection of the variables and parameters defined by the above equations (2) to (7), and in the choice of the forms adopted below for the formula for the characteristic impedance. Except where the contrary is definitely indicated, the formulas that will be adopted for the characteristic impedance are rigorously exact; though the variables F and E are never exactly proportional to the frequency, and the parameters k, g, a, b are never exactly independent of the frequency. If the independant variables were exactly proportional to the frequency and the parameters were exactly independent of the frequency, the graphs of the formulas would by a mere change of scales exactly represent the impedance as an explicit function of the frequency.

With particular regard to considerations (B) and (C) it will be found convenient to divide the further treatment of smooth lines into two main parts, pertaining to open-wire lines and to cables respectively; and then, in each of those parts, to present the impedance formulas in the two forms respectively most suitable for the cases

where the leakance is approximately proportional to the frequency and approximately independent of the frequency, corresponding to consideration (E).

While the classification of smooth lines into open-wire lines and cables is convenient, there is, of course, no very sharp distinction between the open-wire type of lines and the cable type of lines, since the distinction depends on the line parameters and on the frequency range involved, rather than on the physical form of the line; for, any line at sufficiently high frequencies has the open-wire type of characteristics, and at sufficiently low frequencies the cable type of characteristics. With regard to the relative importance of the fundamental line constants R, G, L, C when the frequency range is that of the voice, it may be said that for the open-wire type of lines L and C are of about equal importance, R of secondary, and G of tertiary importance; while, on the other hand, for the cable type, R and C are of about equal importance, L of secondary, and G of tertiary importance. In illustration of the above remarks it may be noted that smoothly loaded cables (unless loaded very lightly) have the open-wire type of characteristics; as have also periodically loaded cables at low frequencies.

Before proceeding to the separate treatments of open-wire lines and cables, it seems desirable to indicate the general nature of the effect produced on the impedance by leakance.

The General Effect of Leakance

The amount of leakance that is normally allowable as regards its attenuating effects is so small as to produce only very slight effects on the characteristic impedance of either type of line (except at very low frequencies).

In ordinary telephone cables the leakance is so small that, except at very low frequencies, the impedance of such cables is very closely the same as in the limiting case of no leakance; whence that limiting case may be taken as being a good approximation to the actual case. In open-wire lines leakance may be much larger than in cables, yet normally it is small enough so that its effects on the impedance are slight, except at very low frequencies, so that usually the limiting value of zero leakance is still a good approximation when calculating the characteristic impedance. However, during wet weather and in particularly humid climates and locations the leakance in open-wire lines becomes large enough to affect the impedance quite appreciably, even within the voice frequency range, while enormously affecting it at very low frequencies.

The general nature of the effect produced on the characteristic impedance K by any value of the leakance G is readily seen from mere inspection of equation (1), so far as regards the absolute value and angle of the impedance. Thus, increasing G from any initial value decreases the absolute value of the impedance and (algebraically) increases the angle. Starting with $G=0$, the angle is negative and has the value $-\frac{1}{2}\tan^{-1}\frac{R}{\omega L}$; increasing G decreases this negative angle until G has become as large as RC/L, when the angle has become zero and the impedance has become equal to the simple value $\sqrt{L/C}$, and thereby equal also to $\sqrt{R/G}$. Increasing G beyond this transition value RC/L toward infinite values gives to the impedance a positive angle which continually increases toward its limiting value $\frac{1}{2}\tan^{-1}\frac{\omega L}{R}$, while the absolute value of the impedance goes on continually decreasing toward its limiting value of zero.

The statements in the foregoing paragraph hold at all frequencies, though the effects of leakance are usually most pronounced at low frequencies. In fact at zero frequency the characteristic impedance of a line having any finite leakance however small is merely $\sqrt{R/G}$; and at frequencies so low that ωL is small compared with R and ωC small compared with G, the impedance K is, approximately,

$$K=\sqrt{\frac{R}{G}}\left\{1+i\omega\left(\frac{L}{2R}-\frac{C}{2G}\right)\right\}$$

and hence is at least roughly equal to $\sqrt{R/G}$.

Of course, with actual lines the whole physically possible range of variation of G from zero to infinite values is never traversed. On the contrary the leakance G even in open-wire lines seldom reaches a value as large as the transition value RC/L and hence the angle seldom becomes positive; while in cables the angle probably always remains negative and indeed is at least roughly equal to its limiting value of $-\frac{1}{2}\tan^{-1}\frac{R}{\omega L}$, except at very low frequencies.

Although, as already indicated, the effects of normal amounts of leakance are usually very small for both cables and open-wire lines, yet the effects in the two cases differ rather markedly in their nature, owing to the difference in the angles of the impedances of these two types of lines; the angle of cables begin almost $-45°$, while that of open-wire lines, though likewise negative, is much smaller (except at very low frequencies).

To formulate analytically the effects of any value of the leakance G, let K_0 denote the value of K when $G=0$, and let ΔK denote the increment $K-K_0$ due to the presence of the leakance G. A suitable measure of the effect of the leakance is then the ratio $\Delta K/K_0$. In order to obtain for the value of this ratio a formula which will be convenient for use with the formula for K (which involves a radical) it is advantageous to write ΔK in the form $\Delta K=(K^2-K_0^2)/(K+K_0)$, and to introduce for brevity the quantity μ defined by the equation

$$\mu=K^2/K_0^2-1. \tag{7.1}$$

This procedure leads readily to the following simple identity for $\Delta K/K_0$, namely

$$\frac{\Delta K}{K_0}=\frac{\mu}{1+\sqrt{1+\mu}}=\sqrt{1+\mu}-1. \tag{7.2}$$

In particular, when this is applied to the formula (1) for K, the value of μ is found to be

$$\mu=\frac{iG/\omega C}{1-iG/\omega C}. \tag{7.3}$$

Equations (7.2) and (7.3) enable the exact value of $\Delta K/K_0$ to be calculated for any value of $G/\omega C$. For small values of $G/\omega C$, the formula for $\Delta K/K_0$ takes the very simple approximate form

$$\Delta K/K_0=iG/2\omega C; \tag{7.4}$$

and this shows that, to the degree of approximation involved, ΔK is proportional to iK_0 through a proportionality factor $(G/2\omega C)$ which is real and positive. Now, for cables, K_0 has an angle of nearly $-45°$; and hence, by (7.4), it is seen that the addition of small leakance increases the resistance component and decreases the negative-reactance component of the impedance by about equal amounts. For open-wire lines, on the other hand, the angle of K_0 is much smaller, though negative, and hence a small increase in the leakance changes the reactance component of the impedance much more than it does the resistance component; evidently, the change in the negative-reactance component is a decrease, but the change in the resistance component may be of either sign, depending on the frequency. This fact regarding the effect of leakance on the resistance component of the impedance is not completely represented by the approximation (7.4)—which indicates the change as being always an increase—but it can be inferred from a study of the exact formulas (7.2) and (7.3). These would have to be employed also if the effects of large leakance were

being studied—or, more generally, large $G/\omega C$. If it were not for a need of these exact formulas, the approximate formula (7.4) would have been derived by the mere application of Taylor's theorem to (1).

Part II

Impedance of Open-Wire Lines

It will be recalled that the characteristic impedance of an ordinary open-wire line depends primarily on its inductance and its capacity, only secondarily on its resistance, and far less still on its leakance; and hence that its impedance is at least roughly equal to $\sqrt{L/C}$.

Of the quantities defined by equations (2), . . . (7), the four most suitable for describing open-wire lines are F, k, b, and a. F is suitable as the independent variable, approximately proportional to the frequency. k is suitable as one parameter. For the other parameter, which evidently must involve the leakance, b or a respectively is the most suitable according as the leakance G is approximately proportional to the frequency or is approximately independent of the frequency. The corresponding suitable forms of the equation for the characteristic impedance K are then

$$K=k\sqrt{\frac{1+iF}{(b+i)F}}, \qquad (8) \qquad\qquad K=k\sqrt{\frac{1+iF}{a+iF}}. \qquad (9)$$

The quantity $k=\sqrt{L/C}$ which occurs in (8) and (9) as a mere factor is significant as being the value that the impedance approaches when the frequency is indefinitely increased[4]; it is also the value the impedance would have at all frequencies if, without changing L and C, the line could be rendered non-dissipative. For ordinary open-wire lines at voice frequencies ($R/\omega L$ small or fairly small compared to unity) it is at least a rough approximation to the value of the impedance. This limiting value $k=\sqrt{L/C}$ will be termed the " nominal impedance " or, more fully, the " nominal characteristic impedance." [5]

The amount $K-k$ by which the characteristic impedance K exceeds the nominal characteristic impedance k will be termed the " excess impedance," and hence its two components the " excess resistance " and the " excess reactance "; (or, more fully, for the three: the " excess characteristic impedance," " excess characteristic resistance," and " excess characteristic reactance," respectively). The latter two

[4] Provided that G/f approaches zero.

[5] Strictly speaking, k varies slightly with the frequency, because of the variations of L and even C.

have respectively the values $M-k$ and N, since k is real; M and N denoting the resistance and reactance components of K. The concept "excess impedance" will be found convenient in various connections, particularly in the description of networks for simulating the impedance of smooth lines.

The ratio K/k of the characteristic impedance K to the nominal impedance k will be termed the "relative impedance" and will be denoted by $z=x+iy$; whence $x=M/k$ will be termed the "relative resistance," and $y=N/k$ the "relative reactance." This complex number z is roughly equal to unity over most of the voice frequency range, and approaches unity as a limit when F is indefinitely increased. Its exact value, written in the two forms corresponding to (8) and (9) respectively, is

$$z=\sqrt{\frac{1+iF}{(b+i)F}}, \qquad (10) \qquad\qquad z=\sqrt{\frac{1+iF}{a+iF}}. \qquad (11)$$

Thus z, which is proportional to the characteristic impedance K (except for the fact that the proportionality factor k is not strictly independent of the frequency), depends merely on the two quantities F and b, or F and a, and hence can be readily represented by tables or graphs.

When z has once been tabulated or graphed the value of K in any specific case (R, G, L, C specified) is readily obtained therefrom by entering such tables or graphs of z with the values of the arguments $F=\omega L/R$ and $b=G/\omega C$ of (10) or the arguments $F=\omega L/R$ and $a=GL/RC$ of (11), and then multiplying the value of z there found by $k=\sqrt{L/C}$. (Graphically this would amount merely to a change of scales if the parameters employed were strictly independent of the frequency.) Thus the function

$$z=\sqrt{(1+iF)/(b+i)F}=\sqrt{(1+iF)/(a+iF)}$$

represents simply and comprehensively the properties of the characteristic impedance of all smooth lines, though it is more suitable for representing open-wire lines than cables.

The two components x and y of z are represented as functions of F by the curves in Figs. 1 and 2 with b and a respectively as parameters. (Explicit formulas for x and y are included in Appendix A.)

The effects produced on $z=x+iy$ by the leakance G are exhibited, in Figs. 1 and 2, through the parameters b and a. These effects may be conveniently represented analytically in a manner formally the same as that already outlined in connection with equations (7.1) and

(7.2); with K, K_0, ΔK (there) corresponding to z, z_0, Δz (here). Thus, by applying (7.1) to (10) and (11),

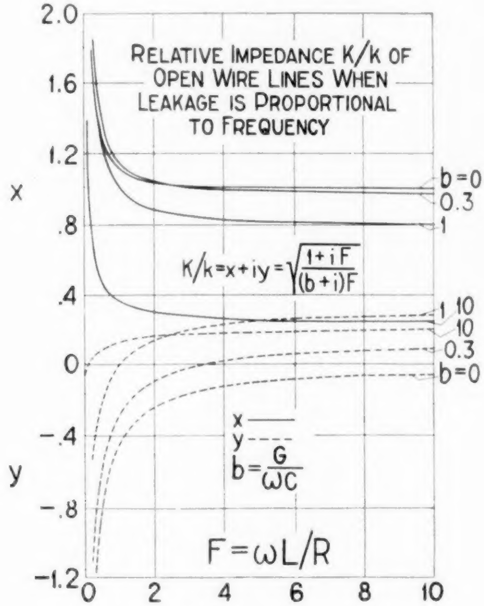


Fig. 1

$$\mu = ib/(1 - ib) = ia/(F - ia).$$

Or, approximately, when b and a are small,

$$\mu = ib = ia/F;$$

and thence, approximately, by (7.2),

$$\Delta z/z_0 = ib/2 = ia/2F.$$

This analysis serves to account, approximately, for the nature of the effects of small leakance, as depicted in Figs. 1 and 2 by the curves for small b and small a. To account for the effects of large leakance, as depicted by the curves for large b and large a, recourse to the exact formula for $\Delta z/z_0$ would be necessary; but the curves for large leakance possess hardly more than academic interest, as will be realized from the remarks already made under the heading *The General Effect of Leakance.*

When there is no leakance ($G=0$, and hence $b=0$ and $a=0$) equations (10) and (11) reduce to the same form, namely

$$z=\sqrt{1-i/F}. \qquad (12)$$

This limiting form of the equation for the relative impedance z is rather important because it is comparatively simple and yet is a close approximation for the impedance of most actual lines except at very low frequencies (since the effects of normal amounts of leakance are

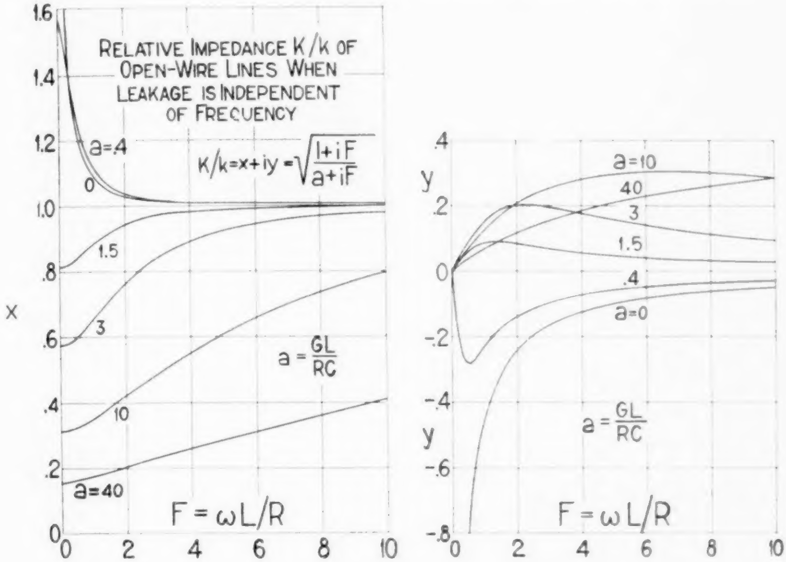


Fig. 2

very small except at very low frequencies). It will therefore now be discussed with some fullness:

For the case of no leakance the formulas for x and y are given under equation (12) in Appendix A; and are graphed in Fig. 1, ($b=0$), and in Fig. 2, ($a=0$). If the wires were devoid of resistance ($R=0$), x would be equal to unity and y would be zero. Thus the effect of wire resistance (in a non-leaky line) is to make x greater than its limiting value unity by the amount $x-1$ (the "relative excess resistance"), and to introduce a negative value of y (the "relative excess reactance," which is equal to the "relative reactance"). Both $x-1$ and $-y$ increase with decreasing F; the increase being slow at large values of F, but more and more rapid as F is decreased. $x-1$

is always smaller than $-y$; and is much smaller except at low values of F, where the two approach equality as F approaches zero. The statements regarding x and y hold also for the effect of wire resistance on the characteristic resistance and the characteristic reactance, since these are (approximately) proportional to x and y respectively, the proportionality factor being the nominal impedance $\sqrt{L/C}$.

Before leaving equation (12) attention will be directed to certain approximate and exact forms of this equation that have been found very useful in devising and proportioning networks for simulating the characteristic impedance of smooth lines, as will appear more fully in the latter part of this paper. At large values of F equation (12) yields immediately the approximation

$$z=1+\frac{1}{8F^2}-i\frac{1}{2F}, \tag{13}$$

whence $x-1$ and y have approximately the values

$$x-1=\frac{1}{8F^2}, \tag{14}$$ $$y=-\frac{1}{2F}. \tag{15}$$

From equation (12) of Appendix A the exact values of $x-1$ and y are known to be

$$x-1=\frac{1}{8F^2}\,\frac{2}{(x+1)x^2}, \tag{16}$$ $$y=-\frac{1}{2Fx}. \tag{17}$$

Thus it is seen that each of the approximations (14) and (15) is always somewhat larger than the exact value, since x is always greater than unity. However, these two approximations are fairly good for values of F as small even as unity, since there x does not exceed 1.1; and they rapidly approach exactness when F is increased, since x rapidly approaches unity. The exact equation for z will now be set down for purposes of reference; by (16) and (17) it is

$$z=1+\frac{1}{8F^2}\,\frac{2}{(x+1)x^2}-i\frac{1}{2Fx}. \tag{18}$$

At small values of F formula (12) shows that z is approximately equal to $z''=x''+iy''$, defined by the equation $z''=1/\sqrt{iF}$. The exact value of the fractional departure $(z-z'')/z''$ is

$$\frac{z-z''}{z''}=\frac{iF}{1+\sqrt{1+iF}}, \tag{18.1}$$

which, at small F, is approximately equal to $iF/2$ merely. Thus, at small F, z exceeds its approximate value z'' by an amount which is

proportional to iz'', through a proportionality factor $(F/2)$ which is real and positive; since the angle of z'' is $-45°$ it follows that x is greater than x'' by about the same amount that $-y$ is less than $-y''$. This analysis serves to account for the shape of the curves of x and y at small values of F and no leakance (the curves $b=0$ in Fig. 1, and $a=0$ in Fig. 2). The shape of the curves at any value of F can be accounted for by means of the exact formula (18.1), or a suitable approximation thereof. In fact formula (18.1) shows immediately that

$$x-x''>(-y'')-(-y)$$

and that this inequality increases with F.

Part III

Impedance of Cables

It will be recalled that the impedance of an ordinary cable depends chiefly on its capacity and resistance, relatively little on its inductance, and far less still on its leakance; and hence that its impedance is at least roughly equal to $\sqrt{R/i\omega C}=(1-i)\sqrt{R/2\omega C}$.

Of the quantities defined by equations (2), . . . (7), the four most suitable for describing cables are E, k, b and g. E is suitable as the independent variable, approximately proportional to the frequency. k is suitable as one parameter. For the other parameter, which evidently must involve the leakance, b or g respectively is the most suitable according as the leakance G is approximately proportional to or approximately independent of the frequency. The corresponding suitable forms of the equation for the impedance are then

$$K=\sqrt{\frac{1+ik^2E}{(b+i)E}}, \qquad (19) \qquad\qquad K=\sqrt{\frac{1+ik^2E}{g^2+iE}}. \qquad (20)$$

These two formulas (19) and (20) for cables are less simple than the corresponding formulas (8) and (9) for open-wire lines, because in (19) and (20) neither of the two parameters enters as a mere factor, and hence the number of effective parameters cannot be reduced to less than two. For purposes of mere specific computations this is not much of a complication; but in graphical representation it is enough to prevent the desired simplicity and compactness, if the representation is required to be exact and comprehensive. (Explicit formulas for the two components M and N of K are included in Appendix A.)

The effects of the leakance G, as represented through the medium of the parameters b and g, may if desired be conveniently formulated and analyzed in a manner formally the same as that already outlined under the heading *The General Effect of Leakance*.

In cables, particularly, the effect of leakance is usually extremely small except at very low frequencies. Hence in the graphical representation of formulas (19) and (20) it will suffice very well to confine ourselves to the limiting case of no leakance ($G=0$, and hence $b=0$ and $g=0$), when these two equations reduce to the same form, namely

$$K=\sqrt{k^2-i/E}. \tag{21}$$

The curves in Fig. 3 represent the resistance and reactance components M and N of K as functions of E with k as parameter.

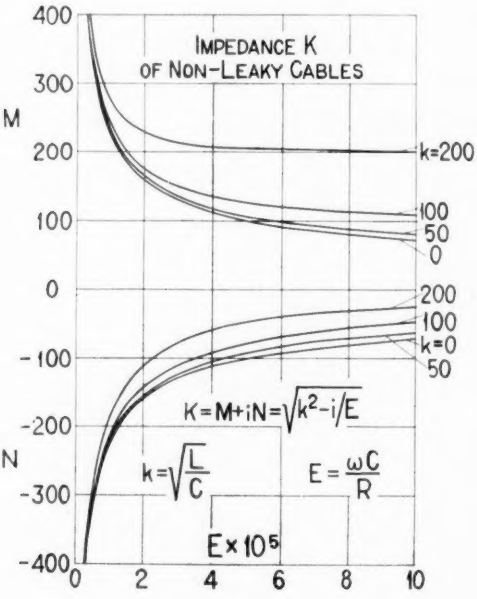


Fig. 3

The effects produced on $K=M+iN$ by the inductance L are exhibited, in Fig. 3, through the parameter k. These effects may be conveniently represented analytically in a manner formally the same as that already employed for the effects of leakance, under the heading *The General Effect of Leakance*. Thus, if K' denotes the value of K,

expressed by (21), when $k=0$, then K' here will correspond to K_0 there; hence $\mu = ik^2E$, and thence, when k^2E is small compared to unity,

$$\Delta K/K' = ik^2E/2.$$

This analysis serves to account approximately for the nature of the effects of small inductance as depicted in Fig. 3. When the leakance is not zero but is small, the effects of inductance are still about the same. The general nature of the effect of the inductance L on the characteristic impedance of any smooth line, so far as regards the absolute value and the angle of the impedance, can be readily determined by mere inspection of equation (1), in a manner similar to that already outlined regarding the effect of leakance under the heading *The General Effect of Leakance.*

An alternative mode of representing the characteristic impedance of cables is suggested by the fact, already mentioned, that the impedance of a cable is at least roughly equal to $\sqrt{R/i\omega C}$, whence its absolute value is at least roughly equal to $\sqrt{R/\omega C}$. This suggests that we study a relative impedance consisting of the ratio of K to $\sqrt{R/\omega_1 C}$, where ω_1 denotes any fixed value of ω; and that we adopt the ratio ω/ω_1 as the independent variable. In this mode of treatment it will be convenient to employ the quantities w, r, F_1, b, b_1 defined by the equations

$$w = \frac{K}{\sqrt{R/\omega_1 C}} = \frac{K}{|K_1|}, \tag{22}$$

$$r = \omega/\omega_1 = f/f_1, \quad (23) \qquad F_1 = \omega_1 L/R, \quad (24)$$

$$b = G/\omega C, \quad (25) \qquad b_1 = G/\omega_1 C. \quad (26)$$

Thus, w denotes the relative impedance to be studied; its real and imaginary components will be denoted by u and v, so that $w = u + iv$. r denotes the relative frequency—relative to any fixed frequency f_1. F_1 is one parameter. The other parameter is, respectively, b or b_1 according as the leakance G is approximately proportional to or approximately independent of the frequency.[6] The corresponding forms of the equation for the relative impedance w are

$$w = \sqrt{\frac{1+iF_1 r}{(b+i)r}}, \quad (27) \qquad w = \sqrt{\frac{1+iF_1 r}{b_1 + ir}}. \quad (28)$$

These are seen to be of the same functional forms as (19) and (20) respectively; with w corresponding to K, r to E, F_1 to k^2, and b_1 to g^2.

[6] It will be noted that b is the same as already defined by (7); b_1 is related to b; and F_1 is related to F, which has already been defined by (2).

In other respects, however, there are marked differences: K is an impedance, while w is a pure number, being the ratio of K to $\sqrt{R/\omega_1 C}$; E, though approximately proportional to the frequency, is not a pure number (for it is not dimensionless), while r is a pure number, being the ratio of the general frequency to any fixed frequency; of the parameters, k^2 is very different from F_1, and b_1 is very different from g^2. The fact that (27) and (28) are of the same functional forms as (19) and (20) respectively renders formally applicable the material pertaining to equations (19) and (20) given in Appendix A.

As already remarked, the effect of leakance in cables is usually extremely small except at very low frequencies. Hence in the graphi-

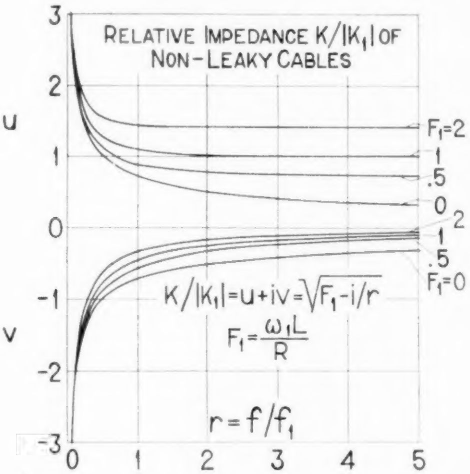


Fig. 4

cal representation of formulas (27) and (28) it will suffice for most purposes to confine ourselves to the limiting case of no leakance ($G=0$, and hence $b=0$ and $b_1=0$), when these two equations reduce to the same form, namely,

$$w=\sqrt{F_1-i/r}. \tag{29}$$

This has the same functional form as (21), with w corresponding to K, r to E, and F_1 to k^2; a circumstance rendering formally applicable the material pertaining to equation (21) given in Appendix A. The curves in Fig. 4 represent the two components u and v of w as functions of r with F_1 as parameter.

Part IV

Networks for Simulating the Impedance of Smooth Lines

Under this heading will be described the various networks devised by the writer, for simulating the characteristic impedance of smooth lines, as mentioned in the latter part of the Introduction. Before proceeding to the systematic description of these networks, some of their practical uses will be mentioned. Foremost of these is their employment for balancing purposes in connection with 22-type repeaters, already spoken of in the Introduction. Another application is for properly terminating an actual telephone line in the field or an artificial line in the laboratory, usually for electrical testing purposes or electrical measurements on the lines. In making certain tests on apparatus normally associated with a telephone line, such line may be conveniently represented for impedance purposes by the appropriate simulating network.

Some of the networks to be shown are potentially equivalent in impedance; but may differ somewhat in cost, space occupied, etc. For the purpose of this paper any two networks will be called "potentially equivalent" if, when the elements of either network are assigned any arbitrary values, the other network can be so proportioned as to have at all frequencies identically the same impedance as the first network. Evidently the mathematical condition for such equivalence is that the expressions for the impedances of the two networks have the same functional forms when the frequency is regarded as the independent variable. The two networks will then have the same number of independent parameters, or degrees of freedom for adjustment; and this number is the same as the minimum number of elements requisite for the construction of a network to have identically the impedance of the given network.

For most of the networks described, there are included design-formulas for the values of the network elements (resistances and capacities). But in any applications requiring the highest simulative precision attainable with such networks, these formulas should be regarded merely as first approximations serving to reduce the requisite detailed design-work down to a relatively small amount but not permitting it to be dispensed with entirely; for the best values of the network elements depend somewhat on the particular frequency-range involved, and on the preassigned weighting of the desired simulative precision with respect to the frequency. Moreover, these formulas completely ignore leakance; while actually leakance may not always be quite negligible, even in the voice frequency-range.

A supplementary semi-graphical method as an aid to finally arriving at the best proportioning of the networks will be found outlined in Appendix C.

The Basic Resistance and the Excess Simulator

The first approximation to a network for simulating the characteristic impedance K of a smooth line is evidently a mere resistance R_1

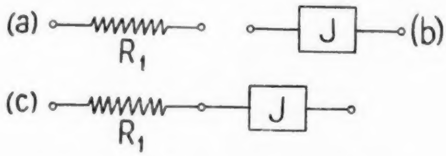


Fig. 5—Synthesis of the General Form of Complete Network. (a). Basic Resistance Element R_1 for Simulating Nominal Impedance. (b). Excess-Simulator J (Abstractly Symbolized) for Simulating Excess Impedance. (c). Complete Network for Simulating Line Impedance

(Fig. 5a) approximately equal to the nominal impedance k of the line, that is,

$$R_1 = \sqrt{L/C}, \tag{30}$$

and this is a very close approximation, for instance, in the case of open-wire lines at the frequencies of carrier current transmission.

Over the voice frequency range, however, a mere resistance does not suffice; since there the excess characteristic impedance $K-k$ is not negligible, particularly at the lower frequencies. But the resistance R_1 equal to the nominal impedance may be retained as the natural basis of a network if it is supplemented by an element or elements such as to approximately simulate the excess characteristic impedance. Such a supplementary network is here termed an "excess-simulator"[7], and is symbolized abstractly by Fig. 5b; while Fig. 5c represents the corresponding complete network consisting of the basic resistance R_1 in series with the excess-simulator, whose impedance is denoted by J. The requisite excess-simulator is obviously less simple in structure and proportioning than the mere basic resistance; whence most of the remainder of this paper will be concerned with various specific types of excess-simulators.

[7] But in practice the term "low frequency corrector" has become rather firmly established. It was suggested by the fact that the excess impedance to be simulated is largest at relatively low frequencies.

The Simplest Excess-Simulator, and Complete Network

The simplest type of excess-simulator is a mere capacity C_1 (Fig. 6b). This is adequate only for those lines whose excess characteristic resistance is negligible; as, for instance, large gauge open-wire lines, and even then not at very low frequencies. The capacity C_1 is cap-

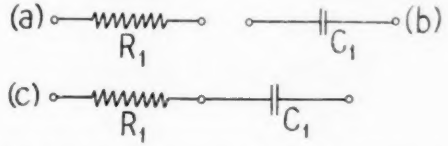


Fig. 6—Synthesis of the Simplest Type of Complete Network. (a). Basic Resistance. (b). Excess-Simulator. (c). Complete Network

able of simulating the reactance X of such a line rather closely, and its proper value for that purpose is approximately

$$C_1 = \frac{2\sqrt{LC}}{R} = C\frac{2\sqrt{L/C}}{R}, \tag{31}$$

although the most suitable value depends somewhat on the specific frequency-range involved. The complete network (Fig. 6c) thus consists merely of a resistance R_1 and a capacity C_1 in series with each other, having approximately the values expressed by (30) and (31).[8]

The simple network in Fig. 6c was devised a good many years ago.[9] The majority of present-day applications require such high simulative precision that the excess characteristic resistance of the line is not negligible, and also a mere capacity does not in all cases simulate the excess characteristic reactance quite as closely as desirable. To meet these needs there have been devised the much more precise, yet fairly simple, excess-simulators and complete networks described under several of the following headings.

Two Precise Types of Excess-Simulators, and Their Limiting Forms

Fig. 7 represents two potentially equivalent[10] excess-simulators that in most cases admit of such proportioning as to simulate with

[8] See Appendix B for the derivation of formula (31) for C_1, and incidentally formula (30) for R_1; and for a discussion of the simulative precision of this network; also for the values R_1' and C_1' requisite for exact simulation at any preassigned single frequency.

[9] In 1913. U. S. Patent No. 1,167,694 of January 11, 1916.

[10] In comparing networks as to equivalence I have found very useful the general theorems on equivalence given by O. J. Zobel in his paper on electric wave-filters in the January Number of this JOURNAL, pages 45–46.

the requisite high precision the excess characteristic impedance of the line; the complete network then consisting of either of these excess-simulators in series with the basic resistance element R_1 of Fig. 5a.

In any specific application the most suitable values for the elements constituting either of the two excess-simulators in Fig. 7 depend

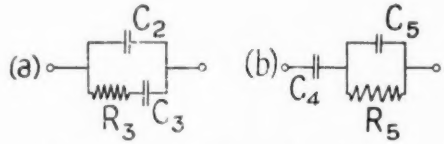


Fig. 7—Two Potentially Equivalent 3-Element Excess-Simulators Possessing High Simulative Precision for Most Applications, Except at very Low Frequencies

somewhat on the particular frequency-range involved, and also on the weighting of the desired simulative-precision with respect to the frequency. As might be expected, therefore, the work of determining closely the best combination of values for the elements of the excess-simulator can hardly avoid a certain amount of tentative detailed design-work; but usually this can be reduced to a relatively small amount by a semi-graphical method such as outlined in the latter part of Appendix C. Moreover, first-approximation values that will usually prove to be rather close, can be quickly found by means of the following approximate design-formulas (32), . . . (37), which are explicit except for containing the single undetermined parameter D. These formulas are such that the excess-simulator will possess high simulative-precision at large and even fairly large values of F, for all physcially admissible values of D $(0 \leq D \leq 1)$; and at the lower values of F will have a considerable range of adjustment by means of D, whose optimum value can be readily determined from inspection of Figs. 8 and 9, as described below. The above-mentioned approximate design-formulas for the elements of the two excess-simulators in Fig. 7 are[11]:

$$C_2 = \frac{2\sqrt{LC}}{R}, \tag{32}$$

$$C_3 = \frac{D}{1-D} \frac{2\sqrt{LC}}{R}, \tag{33}$$

$$R_3 = 2\sqrt{\frac{L}{C}} \tag{34}$$

[11] The first part of Appendix C gives the derivation of these formulas, and also the equation of the curves in Fig. 8.

$$C_4 = \frac{1}{1-D} \frac{2\sqrt{LC}}{R}, \tag{35}$$

$$C_5 = \frac{1}{D} \frac{2\sqrt{LC}}{R}, \tag{36}$$

$$R_5 = D^2 2\sqrt{\frac{L}{C}} \tag{37}$$

When the excess-simulator is proportioned in accordance with these design-formulas the corresponding complete network consisting of such excess-simulator J in series with the basic resistance $R_1 = \sqrt{L/C}$ will possess the simulative precision represented by the set of graphs in Fig. 8, which shows the percentage impedance-departure δ of the

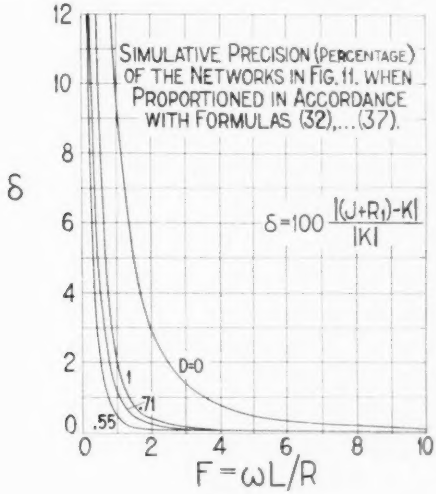


Fig. 8

complete network R_1+J from the line-impedance K, as function of F with D as parameter. In any specific case, where, of course, the F-range would be known, inspection of these graphs (Fig. 8) enables the best value of D to be readily determined, and the corresponding resulting precision δ to be seen as function of F. The curves show that the best value of D is determined by the lowest value of F contemplated, since the departure δ is largest at small values of F and rapidly decreases toward the larger values of F. It will be noted that the curves for the limiting values $D=0$ and $D=1$ have been included

in Fig. 8; the corresponding limiting forms of the excess-simulators are considered a little further on.

Fig. 9, derived from Fig. 8, represents the optimum value of D as function of F; and shows also the corresponding minimum departure δ_m of the complete network. If D is chosen to be the optimum

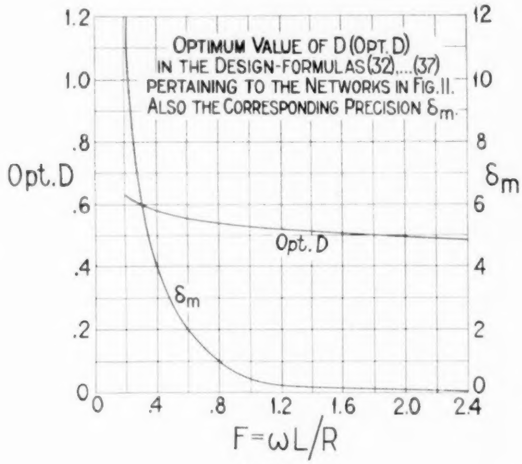


Fig. 9

value at any fixed F, the resulting network will have at that F exactly the departure shown on Fig. 9, but at all other values of F will, of course, have departures larger than those on Fig. 9.

It should be noted that these statements regarding the departures pertain to the network when the excess-simulator is proportioned in accordance with formulas (32), . . . (37). As those are only first-approximation formulas, the ultimate precision attainable will usually be better, and may be adjusted to possess a somewhat different distribution over the frequency-range.

Although the two excess-simulators in Fig. 7 are potentially equivalent as regards impedance there is a slight choice between them from the viewpoints of cost and space occupied. For it is readily seen by mere inspection of the networks at zero frequency that when they have equal impedances the total capacity C_2+C_3 of the excess-simulator in Fig. 7a is equal to merely the capacity C_4 of the excess-simulator in Fig. 7b, thus leaving C_5 in excess. As regards the relative magnitudes of their various elements the two excess-simulators can be

readily compared by means of the following equations (38) derived from (32), . . . (37):

$$\frac{C_4+C_5}{C_2+C_3}=\frac{1}{D}=\frac{C_5}{C_2}=\frac{C_4}{C_3}=\sqrt{\frac{R_3}{R_5}}. \tag{38}$$

Fig. 10 represents the two limiting forms of the excess-simulators in Fig. 7, corresponding to the limiting values 0 and 1 of the parameter D occurring in the design-equations (32), . . . (37). For $D=0$ the limiting form is that in Fig. 10a, and will be recognized as

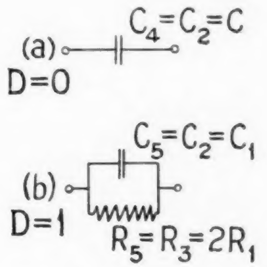


Fig. 10—The Two Limiting Forms of the Excess-Simulators in Fig. 7, Corresponding to the Limits 0 and 1 of the Parameter D

the simple excess-simulator already shown in Fig. 6b consisting of a mere capacity C_1 having the value expressed by (31); while for $D=1$ the limiting form is that in Fig. 10b, and is thus of the same form as one mentioned below, under the heading *Modifications for Very Low Frequencies*, as being capable of furnishing approximate simulation extending down to zero frequency. The departure-curves for these two limiting forms ($D=0$ and $D=1$) are included in Fig. 8, as already mentioned; and from them it is seen that the form in Fig. 10b ($D=1$) possesses much higher simulative precision than the form in Fig. 10a ($D=0$)—as would be expected.

Four Precise Types of Complete Networks, and Their Limiting Forms

Figs. 11a and 11b represent the two potentially equivalent complete networks that can be constructed from the basic resistance R_1 of Fig. 5a, and the excess-simulators in Figs. 7a and 7b respectively; and hence having for their elements approximately the values expressed by equations (30), (32), . . . (37).

Figs. 11c and 11d represent two other complete networks that also are potentially equivalent to those in Figs. 11a and 11b[12].

Appendix D gives the three sets of formulas expressing the values of the elements constituting the networks in Figs. 11b, 11c, 11d re-

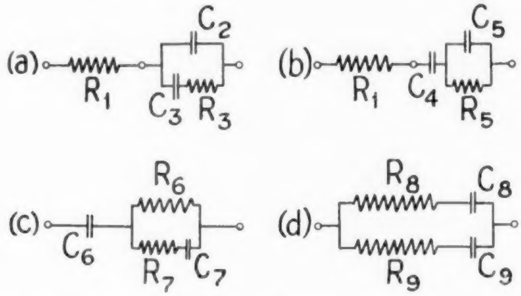


Fig. 11—Four Potentially Equivalent 4-Element Complete Networks Possessing High Simulative Precision for Most Applications, Except at Very Low Frequencies

spectively, in terms of the elements constituting the network in Fig. 11a, when those four networks have equal impedances.

Although the four complete networks in Fig. 11 are potentially equivalent as regards impedance there is some choice among them from the viewpoint of cost and space occupied. For it is readily seen by mere inspection of the networks at zero frequency that, when they have equal impedances,

$$C_2+C_3=C_8+C_9=C_4=C_6. \tag{39}$$

Thus the networks in Figs. 11a and 11d have the same total capacity; and this is less than the total capacity of the network in Fig. 11b by the amount C_5, and is less than the total capacity of that in Fig. 11c by the amount C_7. Similarly by mere inspection of the networks at infinite frequency it is seen that

$$G_6+G_7=G_8+G_9=G_1, \tag{40}$$

the G's being the reciprocals of the R's and thus being the corresponding conductances.

Before leaving Fig. 11 it may be noted that the network in Fig. 11d has the same form as though obtained by connecting in parallel two networks having the same form as Fig. 6c but with elements R_1', C_1' and R_1'', C_1'', say. Now it is known that, in most applications, the

[12] In connection with Figs. 11b and 11c the network shown in U. S. Patent No. 1,240,213 of September 18, 1917 may be of some interest.

network in Fig. 11d has much higher simulative precision than that in Fig. 6c. These considerations suggest the possibility of attaining still higher precision by connecting in parallel several such networks, having constants R_1', C_1'; R_1'', C_1''; R_1''', C_1''', ...

Fig. 12 represents the two limiting forms of the network in Fig. 11, corresponding to the limiting values 0 and 1 of the parameter D occurring in the design-equations (32), . . . (37). For $D=0$ the limiting form is that represented in Fig. 12a, and this will be recognized

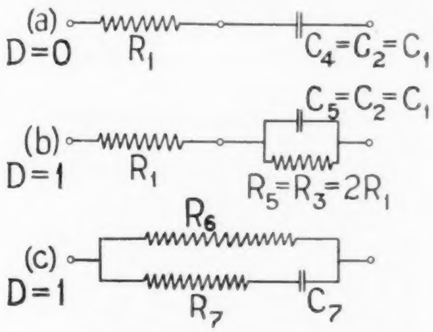


Fig. 12—The Two Limiting Forms of the Networks in Fig. 11, Corresponding to the Limits 0 and 1 of the Parameter D. Networks (b) and (c) are Potentially Equivalent

as the simple 2-element network already shown in Fig. 6c; while for $D=1$ the limiting forms are the two potentially equivalent 3-element networks represented in Figs. 12b and 12c, and are thus of the same forms as two mentioned below, under the heading *Modifications for Very Low Frequencies,* as being capable of furnishing approximate simulation extending down to zero frequency. The values of the elements of the network in Fig. 12c in terms of the elements of the network in Fig. 12b, for equivalence of these two networks as regards impedance, are

$$R_6 = R_1 + R_3 = 3R_1, \tag{41}$$

$$R_7 = R_1(1 + R_1/R_3) = 3R_1/2, \tag{42}$$

$$C_7 = \frac{C_2}{(1+R_1/R_3)^2} = \frac{4C_2}{9}. \tag{43}$$

Thus the network in Fig. 12c requires only four-ninths as much capacity as the network in Fig. 12b.

Modifications for Very Low Frequencies

Thus far the present paper has dealt with the characteristic impedance of smooth lines as distinguushed from their sending-end impedance, strictly speaking. The two are closely equal when the lines are electrically long, which is usually the case for the telephonic frequency range; but at very low frequencies the sending-end impedance of even a rather long line may depend very greatly on the distant terminating impedance and hence depart widely from the characteristic impedance. In case the terminating impedance is conductive to direct current the sending-end impedance of even a strictly non-leaky line would have a finite value at zero frequency; its resistance component evidently being equal to the total line-wire resistance plus the terminating resistance, while its reactance component would, of course, be zero. Actually, on account of line leakance, the resistance component would be somewhat less; and in case the distant terminating impedance permits no passage of direct current the sending-end impedance of the line at zero frequency would depend largely on the line leakance.

Most of the simulating networks thus far described were devised primarily with regard to the voice range of frequencies, without reference to frequencies very far below that range. At very low frequencies these networks become unsuitable because their impedance is not only much too large but also has not even approximately the proper angle. There have not been many occasions for modifying the networks so as to extend their range of simulation down toward zero frequency; but it seems likely that in most cases the requisite modification in the network impedance could be attained, at least roughly, by shunting the excess-simulator (Fig. 5b) with a mere resistance S' approximately equal to the zero-frequency sending-end resistance of the line diminished by the resistance R_1 of the basic resistance element. Clearly this modification will give the network the desired impedance at zero-frequency, without affecting its impedance at infinite frequencies; since the impedance of the unshunted excess-simulator is infinite [13] at zero-frequency and is zero at infinite frequencies. At the intermediate frequencies the resulting modification would doubtless be slight except toward the lower frequencies, where it would increase more and more rapidly as zero-frequency is approached. Of course, the addition of the modifying element S' would usually entail some alterations in the proportioning of the

[13] Except for the limiting form in Fig. 10b.

original network, as indicated by Fig. 13a, where the altered values of J and R_1 are denoted by J' and R_1' respectively.

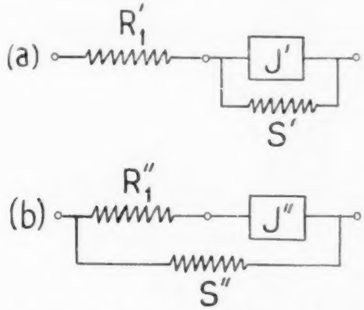


Fig. 13—Two Potentially Equivalent Modifications for Extending Range of Simulation Down to Zero Frequency. (a). Modification by Shunting the Excess-Simulator J'. (b). Modification by Shunting the Complete Network $R_1''+J''$

Fig. 13b represents an alternative but potentially equivalent form of modification, obtained by shunting the original form of network (Fig. 5c) with a resistance S''; and the conditions for equivalence are

$$S''=S'+R_1', \tag{44}$$

$$R_1''=R_1'(1+R_1'/S'), \tag{45}$$

$$J''=J'(1+R_1'/S')^2. \tag{46}$$

Since the shunts S' and S'' are potentially equivalent in their effects their simultaneous application would be potentially equivalent to the application of either alone.

Thus far the suggested modifications have been stated only with reference to the excess-simulator regarded abstractly. When the specific structure of the excess-simulator is regarded, the modifications can take several different forms which, for any one excess-simulator, are equivalent as regards impedance. Certain of these are noted in the following paragraphs:

Among the modified excess-simulators will evidently be found one having the limiting form already depicted in Fig. 10b.

Fig. 14 represents by (a) and (b), respectively, the 3-element excess-simulators in Figs. 7a and 7b modified by the shunt resistance S', and thereby converted to 4-element excess-simulators. Figs. 14c and 14d represent two other 4-element excess-simulators that are potentially equivalent to those in Figs. 14a and 14b as regards impedance.

Before leaving Fig. 14 it may be noted that the excess-simulator in Fig. 14d has the same form as though obtained by connecting in series two of the simple form of modified excess-simulator in Fig. 10b having elements C_2', R_3' and C_2'', R_3'', say. This observation suggests

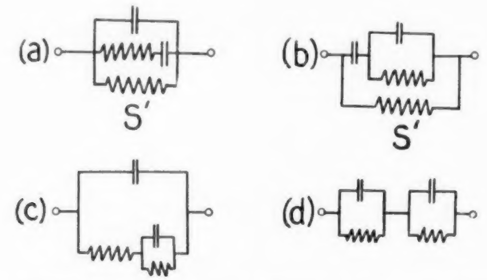


Fig. 14—Four Potentially Equivalent 4-Element Excess-Simulators Embodying Shunt-Resistance Modifiers for Extending the Range of Simulation down to Zero Frequency

the possibility of attaining still higher simulative precision for a modified excess-simulator by connecting in series several such simple modified excess-simulators.

Among the modified complete networks will evidently be found two having respectively the forms already depicted in Figs. 12b and 12c.

Fig. 15 represents four potentially equivalent complete networks derived from Fig. 11d by application of a shunt resistance S''. The

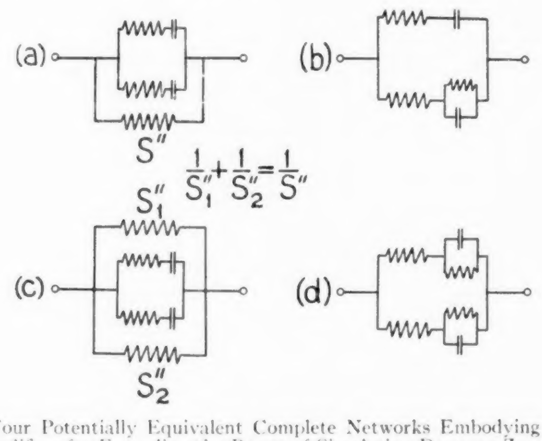


Fig. 15—Four Potentially Equivalent Complete Networks Embodying Shunt-Resistance Modifiers for Extending the Range of Simulation Down to Zero Frequency

forms in Figs. 15c and 15d, though each containing a superfluous element, are of interest because they have the same forms as though obtained by connecting in parallel two networks of the forms already depicted in Figs. 12c and 12b respectively.

Modifications For Leaky Lines

For lines whose leakance is not quite negligible a study of the formulas and graphs of the line impedance indicates that the effects of such leakance can be sufficiently taken into account by a mere slight reproportioning of the network without the addition of any further element, except that a small series inductance might be a slight improvement in those cases where the leakance increases rapidly with the frequency.

APPENDIX A

EQUATIONS OF THE COMPONENTS OF THE LINE IMPEDANCE

This Appendix contains the equations for the rectangular components and the equation for the angle of the relative impedance z and of the impedance K, corresponding to most of the various forms of the equations employed in this paper for expressing z and K. It thereby includes the equations for all the graphs employed in representing z and K. It contains also the equations for the network of curves of z and K in the complex plane for certain of the more important limiting cases involving not more than one parameter.

With regard to the notation it will be recalled that $z=x+iy$ and $K=M+iN$. The angles of z and K will be denoted by $ag\ z$ and $ag\ K$, respectively, "ag" being an abbreviation for "angle of".

Equation (10):

$$z=\sqrt{\frac{1+iF}{(b+i)F}},$$

$$x=\sqrt{\frac{b+F+\sqrt{(1+b^2)(1+F^2)}}{2(1+b^2)F}},$$

$$y=-\frac{1-bF}{2(1+b^2)Fx},$$

$$ag\,z=-\tfrac{1}{2}\tan^{-1}\frac{1-bF}{b+F}.$$

Equation (11):

$$z=\sqrt{\frac{1+iF}{a+iF}};$$

$$x=\sqrt{\frac{a+F^2+\sqrt{(1+F^2)(a^2+F^2)}}{2(a^2+F^2)}},$$

$$y=-\frac{(1-a)F}{2(a^2+F^2)x},$$

$$ag\,z=-\tfrac{1}{2}\tan^{-1}\frac{(1-a)F}{a+F^2}.$$

Equation (12):

$$z=\sqrt{1-i/F};$$

$$x=\frac{1}{\sqrt{2}}\sqrt{1+\sqrt{1+1/F^2}},$$

$$y=-1/2Fx=-\sqrt{x^2-1},$$

$$x-1=\frac{1}{8F^2}\,\frac{2}{(x+1)x^2},$$

$$ag\,z=-\tfrac{1}{2}\cot^{-1}F.$$

The relation $y=-\sqrt{x^2-1}$ can be written in the form

$$\frac{x-1}{-y}=\sqrt{\frac{x-1}{x+1}},$$

which shows that $x-1$ is always smaller than $-y$; and is very much smaller except at small values of F, where the two approach equality as F approaches zero.

The locus of z in the xy-plane is the hyperbola $x^2-y^2=1$. For any preassigned value of F the corresponding values of x and y on this locus can be accurately calculated by means of the above equations for x and y. For any pair of values of x and y situated on this locus the corresponding value of F is given by $F=-1/2xy$.

Equation (19):

$$K=\sqrt{\frac{1+ik^2E}{(b+i)E}};$$

$$M=\sqrt{\frac{b+k^2E+\sqrt{(1+b^2)\,(1+k^4E^2)}}{2(1+b^2)E}},$$

$$N=-\frac{1-bk^2E}{2(1+b^2)EM},$$

$$ag\,K=-\tfrac{1}{2}\tan^{-1}\frac{1-bk^2E}{b+k^2E}.$$

Equation (20)

$$K=\sqrt{\frac{1+ik^2E}{g^2+iE}};$$

$$M=\sqrt{\frac{g^2+k^2E^2+\sqrt{(1+k^4E^2)(g^4+E^2)}}{2(g^4+E^2)}},$$

$$N=-\frac{(1-g^2k^2)E}{2(g^4+E^2)M},$$

$$ag\,K=-\tfrac{1}{2}\tan^{-1}\frac{(1-g^2k^2)E}{g^2+k^2E^2}.$$

Equation (21):

$$K=\sqrt{k^2-i/E};$$

$$M=\frac{1}{\sqrt{2}}\sqrt{k^2+\sqrt{k^4+1/E^2}},$$

$$N=-1/2EM=-\sqrt{M^2-k^2},$$

$$M-k=\frac{1}{8E^2}\,\frac{2}{(M+k)M^2},$$

$$ag\,K=-\tfrac{1}{2}\cot^{-1}k^2E.$$

The relation $N=-\sqrt{M^2-k^2}$ can be written in the form

$$\frac{M-k}{-N}=\sqrt{\frac{M-k}{M+k}},$$

which shows that $M-k$ is always smaller than $-N$, though the two approach equality when E approaches zero.

The network of curves of K in the MN-plane are the equi-k curves consisting of the family of hyperbolas $M^2-N^2=k^2$, and the equi-E curves consisting of the family of hyperbolas $MN=-1/2E$.

Appendix B

On the Simple Type of Complete Network (Fig. 6)

The network in Fig. 6c consisting of a resistance R_1 and capacity C_1 in series with each other and having the values expressed by equations (30) and (31) was originally arrived at by working with values of F large or at least fairly large compared with unity; for then, by equation (12), the characteristic impedance K has approximately the value.

$$K=k-ik/2F. \qquad (1\text{-B})$$

This represents K as having a resistance component k that is independent of frequency, and a reactance component $-k/2F$ that is negative and inversely proportional to the frequency f (since $F=\omega L/R$) and thus leads exactly to the values of R_1 and C_1 expressed by (30) and (31), whence the impedance of this network is exactly equal to the approximate value of the line impedance expressed by (1-B).

To obtain more precise and comprehensive knowledge regarding the simulative precision of this network its exact impedance $k-ik/2F$ will here be compared with the exact value of the line impedance (when leakance is neglected). For this purpose it is convenient to employ the line impedance in the form

$$K=xk-ik/2xF, \tag{2-B}$$

obtained by means of the relation $y=-1/2Fx$ found under equation (12) in Appendix A. The equation (2-B) shows that to exactly simulate the line impedance by a resistance R_1' and capacity C_1' in series with each other these would have to possess the values

$$R_1'=x\sqrt{L/C}, \tag{3-B}$$

$$C_1'=\frac{2x\sqrt{LC}}{R}, \tag{4-B}$$

which differ only by the factor x from the values of R_1 and C_1 expressed by (30) and (31). Thus the ideal resistance R_1' and capacity C_1' for exactly simulating the line impedance would vary with F in precisely the same way as x varies with F. Moreover the ratio of these ideal values to the fixed values of R_1 and C_1 expressed by (30) and (31) is merely x. By reference to Fig. 1 (with $b=0$) it will be seen that, except at small values of F, the factor x is nearly independent of F and is only slightly greater than unity. Thus the values of R_1 and C_1 determined by means of equations (30) and (31) are slightly too small at all frequencies; while the values determined by means of equations (3-B) and (4-B), for any specified frequency (by inserting the appropriate value of x), are slightly too small at lower frequencies and slightly too large at higher frequencies.

Since (3-B) can, by (30), be written in the form

$$R_1'=R_1+(x-1)\sqrt{L/C}, \tag{5-B}$$

and since x is always greater than unity, it is seen that the simulation can be somewhat improved by supplementing the excess-simulator with a small series resistance element R_{11}, the ideal value of which would be

$$R_{11}=(x-1)\sqrt{L/C}. \tag{6-B}$$

Actually, since x varies with frequency, R_{11} is limited to some compromise value. In practice R_{11} would usually be combined with the basic resistance R_1, though the functions of the two are distinctly different. (If the requisite value of R_{11} were negative, R_1 would merely be decreased by that amount.)

Appendix C

On the Precise Types of Excess-Simulators (Fig. 7)

The two sets of formulas (32), (33), (34) and (35), (36), (37), representing first-approximations to the proper values of the elements constituting the excess-simulators in Figs. 7a and 7b respectively, were originally obtained by working with values of F large or at least fairly large compared with unity; for then, by (13), the excess characteristic impedance $K-k$ has approximately the value

$$K-k=\frac{k}{8F^2}-i\frac{k}{2F}, \tag{1-C}$$

while, at large or fairly large values of T, the impedance $J=P+iQ$ of each excess-simulator in Fig. 7 can be expressed approximately by the equation

$$J=\frac{P_0}{T^2}-i\frac{(1+t)P_0}{T}, \tag{2-C}$$

derived from the exact equation (16-C) below, in which t, P_0, and T have the values defined by the following two sets of equations (3-C), (4-C), (5-C) and (6-C), (7-C), (8-C) for the excess-simulators in Figs. 7a and 7b respectively:

$$t=C_2/C_3, \tag{3-C}$$

$$P_0=\frac{R_3}{(1+t)^2}, \tag{4-C}$$

$$T=\frac{\omega C_2 R_3}{1+t}, \tag{5-C}$$

$$t=C_5/C_4, \tag{6-C}$$

$$P_0=R_5, \tag{7-C}$$

$$T=\omega C_5 R_5, \tag{8-C}$$

P_0 thus being the value of P at $\omega=0$. Comparison of the approximate equations (1-C) and (2-C) gives immediately

$$P_0/T^2=k/8F^2, \tag{9-C}$$

$$(1+t)P_0/T=k/2F, \tag{10-C}$$

as the two conditions that are necessary and sufficient for (approximate) equality of J and $K-k$ at large values of F and T. This pair

of equations is equivalent to the more convenient equations (11-C),

$$\frac{T}{F}=\frac{4}{1+t}=\sqrt{\frac{8P_0}{k}}. \qquad (11\text{-C})$$

Thus the ratio of T to F is fixed as soon as either t or P_0/k is fixed. It will be convenient to adopt $\sqrt{P_0/2k}$ as the arbitrary quantity and to denote it by D, so that

$$D=\sqrt{P_0/2k}, \qquad (12\text{-C})$$

whence

$$P_0=2D^2k, \qquad (13\text{-C})$$

and

$$t=\frac{1}{D}-1, \qquad (14\text{-C})$$

and

$$T=4DF. \qquad (15\text{-C})$$

Since only positive values of t and D are physically admissible, equation (14-C) shows that the admissible range of D is 0 to 1.

From (13-C), (14-C), (15-C) and the defining equation $F=\omega L/R$ the two sets of equations (32), (33), (34) and (35), (36), (37) follow readily from the two sets of defining equations (3-C), (4-C), (5-C) and (6-C), (7-C), (8-C), respectively.

The formula for plotting the curves in Fig. 8 depends on the exact equation for J/P_0 which is

$$\frac{J}{P_0}=\frac{1}{1+T^2}-i\frac{t+(1+t)T^2}{T(1+T^2)}. \qquad (16\text{-C})$$

By substituting herein the values of P_0, t, and T expressed by (13-C), (14-C), (15-C) the equation for J/k becomes

$$\frac{J}{k}=\frac{2D^2}{1+16D^2F^2}-i\frac{1+16D^2F^2-D}{2F(1+16D^2F^2)}, \qquad (17\text{-C})$$

which is thus the exact formula for the relative impedance J/k of each of the excess-simulators in Fig. 7 when these are proportioned in accordance with the formulas (32), . . . (37).

A semi-graphical method will now be outlined in the remainder of this Appendix. In this method the ratio T/f is of frequent occurrence and will be denoted by d. Then, recalling that $P+iQ=J$, it will be seen from equation (16-C) that P/P_0 depends only on f and d; while Q/P_0 depends on f, d, and t. These observations are the basis for the method now to be described for evaluating the three para-

meters P_0, d, and t which implicitly determine the elements of the excess-simulators in Fig. 7.

In the first step of this method the two parameters d and P_0 are so chosen that the resistance component P of the excess-simulator will be approximately equal to the excess resistance $M-k$ of the line-impedance K, over the specific f-range contemplated, or else will differ therefrom by a nearly constant amount, which can be approximately simulated by a mere series resistance element. In the second step of the method the remaining parameter, t, is so chosen that the reactance component Q of the excess-simulator will be approximately equal to the reactance N of the line impedance, when d and P_0 have the pair of values already chosen in the first step. The technical procedure in these two steps may now be formulated explicitly as follows:

First, over the contemplated f-range, plot a set of curves representing P/P_0 as function of f with d as parameter; and on the same sheet a set of curves representing $(M-k)/P_0$ as function of f with P_0 as parameter. To evaluate d and P_0 choose (by interpolation, if necessary) such P/P_0-curve and $(M-k)/P_0$-curve as most closely coincide. A preliminary idea regarding the useful ranges of d and P_0 can be readily obtained from the approximate formulas (15-C) and (13-C), together with Fig. 8.

Second, on another sheet plot as function of f that particular N/P_0-curve having as parameter the value of P_0 already found in the first step. With this value of P_0 and the corresponding value of d, as found in the first step, plot also a sufficient set of Q/P_0-curves as function of f with t as parameter to find the one that coincides most closely with the single N/P_0-curve already plotted. To abridge this step tentative values for t can be readily obtained from the approximate formula (14-C), together with Fig. 8. But the useful range of t can be demarcated more closely by solving for t the equation obtained by equating the expressions for Q and N; the value for t thus found is [14]

$$t=-\frac{dfN}{P_0}-\frac{d^2f^2}{1+d^2f^2}.$$

This may even be plotted, as function of f, to see whether the requisite value of t varies much in the contemplated f-range.

If the best compromise value of t found in the second step is unsatisfactory as regards simulation of N by Q, it will be necessary to revert to the first step, choose some other pair of values for d and

[14] It will be recalled that the line-reactance N is practically always negative.

P_0, and with these repeat the second step. In this connection it should be noted that, in the first step, it is not necessary to choose the P/P_0-curve and $(M-k)/P_0$-curve which most closely coincide; on the contrary it suffices to choose two curves that are closely parallel (that is, have closely equal slopes at each f). For, corresponding to the nearly constant distance between such two curves, it will only be necessary to supplement the excess-simulator with a series resistance element R_{11}—which will thus in the complete network be also in series relation to the basic resistance R_1 and hence can be merged therewith (even when the requisite R_{11} is negative, provided it is less than R_1 in absolute value).

After the parameters t, P_0, and $d=T/f$ have been evaluated, the values for the elements of the excess-simulators in Figs. 7a and 7b can be readily obtained from the two sets of equations (3-C), (4-C), (5-C) and (6-C), (7-C), (8-C), respectively; it thus being found that

$$C_2=d/2\pi(1+t)P_0,$$
$$C_3=d/2\pi(1+t)tP_0,$$
$$R_3=P_0(1+t)^2,$$

$$C_4=d/2\pi tP_0,$$
$$C_5=d/2\pi P_0,$$
$$R_5=P_0.$$

The requisite value for the supplementary series resistance element R_{11} is evidently

$$R_{11}=P_0\left(\frac{M-k}{P_0}-\frac{P}{P_0}\right),$$

which will be approximately independent of f if the curves of P/P_0 and $(M-k)/P_0$ chosen in the first step are approximately parallel. If the requisite value of R_{11} is negative, the basic resistance R_1 will merely be decreased by that amount.

For the limiting form of excess-simulator in Fig. 10b the design-procedure is considerably simpler, because the parameter t is fixed $(t=0)$. The two remaining parameters d and P_0 can be evaluated by inspection of two sheets of curves plotted as functions of f: One sheet containing a set of curves of P/P_0 with d as parameter, and curves of $(M-k)/P_0$ with P_0 as parameter; and the other sheet, curves of Q/P_0 with d as parameter, and curves of N/P_0 with P_0 as parameter.

Instead of f as the independent variable it may be more convenient to employ some quantity proportional to f (for instance, F or E); likewise, instead of P, P_0, Q, M, N, some quantities proportional to them (for instance, their ratios to k).

Appendix D

Relative Values of the Elements in the Four Precise Types of Complete Networks (Fig. 11)

The following three sets of formulas express the values of the elements constituting the networks in Figs. 11b, 11c, 11d, respectively, in terms of the elements constituting the network in Fig. 11a when those four networks have equal impedances. These formulas involve the two ratios ξ and ζ pertaining to the network in Fig. 11a and defined by the equations

$$\xi = C_3/C_2, \quad \zeta = R_1/R_3.$$

For Fig. 11b,

$$\frac{R_5}{R_3} = \left(\frac{\xi}{1+\xi}\right)^2, \qquad \frac{C_4}{C_3} = \frac{1+\xi}{\xi}, \qquad \frac{C_5}{C_3} = \frac{1+\xi}{\xi^2}.$$

For Fig. 11c,

$$\frac{R_6}{R_3} = \zeta + \left(\frac{\xi}{1+\xi}\right)^2, \qquad \frac{C_6}{C_3} = \frac{1+\xi}{\xi},$$

$$\frac{R_7}{R_3} = \zeta + \left(\zeta\frac{1+\xi}{\xi}\right)^2, \qquad \frac{C_7}{C_3} = \frac{1}{(1+\xi)\left(\frac{\xi}{1+\xi} + \zeta\frac{1+\xi}{\xi}\right)}.$$

For Fig. 11d,

$$\frac{R_8}{R_3} = \frac{2\zeta\tau}{(\eta+\tau) - 2(1+\xi)},$$

$$\frac{R_9}{R_3} = \frac{2\zeta\tau}{2(1+\xi) - (\eta-\tau)},$$

$$\frac{C_8}{C_3} = \frac{2\xi - \zeta(1+\xi)(\eta-\tau)}{2\xi\zeta\tau},$$

$$\frac{C_9}{C_3} = \frac{\zeta(1+\xi)(\eta+\tau) - 2\xi}{2\xi\zeta\tau},$$

where $\eta = 1 + \xi + \frac{\xi}{\zeta}$, $\quad \tau = \sqrt{\left(1+\xi+\frac{\xi}{\zeta}\right)^2 - 4\frac{\xi}{\zeta}}$.

Appendix E

Illustrative Example

The example contained in this Appendix serves two purposes. First, it illustrates the use of two types of the general line-impedance graphs contained in Parts II and III of the paper; second, it illustrates the first-approximation design of a simulating network by means of the method in Part IV.

The specific example chosen pertains to a well-insulated open-wire line consisting of two horizontal parallel copper wires, of No. 12 N. B. S. gauge, having per loop-mile the constants

$R = 10.4$ ohms, $L = .00367$ henry, $C = .00835 \times 10^{-6}$ farad,

the leakance G being regarded as negligible. This particular type and gauge of line was chosen because it is rather extensively employed in practice, and also because its excess impedance is far from being negligible even as regards its resistance component.

For the illustratve purposes contemplated, it will be supposed that it is desired to evaluate the resistance and reactance components M and N of the characteristic impedance K of this line over the frequency-range from 200 to 2500 cycles per second; and also to design a network for approximately simulating this impedance over that frequency-range, and to determine the simulative precision of such network.

The procedure and results are indicated by the following table together with the supplementary description coming thereafter.

f	F	x	$-y$	M	$-N$	r	u	$-v$	δ_0	δ
200	.444	1.32	.86	876	570	.222	1.87	1.22	——	3.0
300	.666	1.19	.64	790	425	.333	1.69	.91	16.6	1.3
500	1.110	1.08	.42	717	279	.555	1.53	.60	8.1	3
800	1.776	1.04	.27	691	179	.888	1.48	.38	3.6	1
1200	2.664	1.02	.19	676	126	1.332	1.45	.27	1.7	1
1600	3.552	1.01	.14	670	93	1.776	1.43	.20	1.0	1
2000	4.440	1.01	.12	670	80	2.220	1.43	.17	.6	1
2300	5.106	1.01	.10	670	66	2.553	1.43	.14	.5	1
2500	5.550	1.01	.10	670	66	2.775	1.43	.14	.4	1
∞	∞	1	0	663	0	∞	$\sqrt{2}$	0	0	0

The first column in the table is a set of values of the frequency f distributed over the specified range 200 to 2500.

The columns headed $F, x, -y, M, -N$, show the successive steps in evaluating the characteristic impedance $K = M + iN$ by means of

Fig. 2[15], with $a=0$ (since the leakance G is neglected). The F-column was obtained from the f-column by means of the equation $F=\omega L/R=.00222f$. Next the values of x and y were read from the curves $a=0$ in Fig. 2. Finally M and N were obtained from $M=kx$ and $N=ky$, with $k=\sqrt{L/C}=663$ ohms (whence M and N are in ohms). From the table it will be noted that at $f=200$ the excess resistance is about one-third as large as the nominal impedance, and the reactance is about nine-tenths as large as the nominal impedance.

The columns r, u, $-v$ show the steps in evaluating M and N by means of Fig. 4. After choosing F_1 (in Fig. 4) equal to 2[16], the r-column was obtained from the f-column by means of the equation $r=f/f_1=2\pi Lf/RF_1=.00111f$. Next the values of u and v were read off from the curves $F_1=2$ in Fig. 4. Finally the values of M and N were obtained from $M=|K_1|u$ and $N=|K_1|v$, with $|K_1|=\sqrt{R/\omega_1 C}=\sqrt{L/CF_1}=469$.

The two last columns (δ_0, δ) of the table show the percentage precision of certain simulating networks designed in accordance with the first-approximation methods of this paper, and having for their elements the values given in the last paragraph of this Appendix.

δ_0 pertains to the simple 2-element network in Fig. 6c when proportioned in accordance with equations (30) and (31). The values of δ_0 were read from the curve $D=0$ in Fig. 8. The precision at the lower frequencies could be considerably improved by the addition of a small resistance R_{11} (as already noted in connection with equation (6-B) of Appendix B), but with a corresponding sacrifice in the rest of the range.

δ pertains to the 4-element networks in Figs. 11a and 11b when proportioned in accordance with (30), (32), (33), (34) and (30), (35), (36), (37) respectively; and δ pertains also to the networks in Figs. 11c and 11d when these are proportioned, by Appendix D, so as to be equivalent to the network in Fig. 11a. The values of δ were read from the curve $D=0.55$ in Fig. 8, this value of D being known from Fig. 9 to be about the best.

The values of the elements of the networks to which δ_0 and δ pertain will now be set down, each preceded by a reference to the corresponding diagram and design-formulas:

Fig. 6c—Formulas (30) (31); Precision δ_0.

$R_1=663$ ohms, $\qquad C_1=1{,}063\text{x}10^{-6}$ farad.

[15] Or Fig. 1, with $b=0$; but Fig. 2 is plotted to a larger scale.

[16] $F_1=1$ would be somewhat preferable for reading off values; but the principles are more clearly exhibited by choosing for F_1 some other value than unity.

Fig. 11a—Formulas (30), (32), (33), (34); Precision δ.

$R_1 = 663$ ohms, $C_2 = 1.063 \times 10^{-6}$ farad, $C_3 = 1.300 \times 10^{-6}$ farad, $R_3 = 1326$ ohms.

Fig. 11b—Formulas (30), (35), (36), (37); Precision δ.

$R_1 = 663$ ohms, $C_4 = 2.362 \times 10^{-6}$ farad, $C_5 = 1.932 \times 10^{-6}$ farad, $R_5 = 401$ ohms.

NOTE:—To furnish sufficiently high precision for most engineering applications the curves and the cross-section lines in the line-impedance charts would evidently have to be drawn at much closer intervals than has been done in the present paper, where the purpose of the charts is mainly qualitative, or only roughly quantitative, to exhibit the general nature of the functions involved.

Practical Application of Carrier Telephone and Telegraph in the Bell System

By ARTHUR F. ROSE

IN 1918 it was announced that the engineers of the Bell System had perfected carrier current telephone apparatus to such a point that four talking circuits had been added to one pair of wires already in use for telephone and telegraph communication and were being used commercially between Pittsburgh and Baltimore for providing needed telephone facilities. Since that time the growth of carrier application in the Bell System has been quite rapid. The purpose of this paper is to summarize the applications of carrier up to the present time and give a few typical examples where it has been found economical to provide circuits by means of carrier rather than by other types of facilities.

Principles of Operation

The theory of carrier current systems, together with a historical sketch, was presented by Messrs. Colpitts and Blackwell before the American Institute of Electrical Engineers in February, 1921, and was published in Volume XL of the Transactions of the Institute. For those who do not wish to go into the detailed theory given in that paper, it may suffice to say that in a carrier current system a number of telephone or telegraph messages are simultaneously superposed on a single pair of wires by means of high frequency currents of different frequencies on which the individual messages are impressed. It is from this principle that the carrier current systems get their name, as the individual high frequency currents may be said to "carry" the telegraph or telephone messages. By using different frequencies for the carrier currents, the individual messages retain distinctive features which enable them to be separated one from another at the receiving end of the circuit.

On account of the much higher frequencies that are used in carrier operation, the carrier currents are attenuated more rapidly than the ordinary low frequency voice currents. This requires that repeaters be located at frequent intervals in a carrier system. In these repeaters all the carrier channels are amplified together although the ordinary voice frequency channel is separated out and amplified in its own repeater.

The telephone and telegraph carrier systems although alike in their essentials differ very materially in the details of their operation. With the present equipment the frequencies employed in carrier telephony are much higher than in carrier telegraphy, thereby requiring more frequent repeater stations. In both telephone and telegraph systems it is necessary to provide for two way operation. This may be accomplished by using different carrier frequencies in the two directions or by using the same frequency in each direction with directional selectivity obtained by the three-winding coil (hybrid coil) used in repeater work. In this latter case it is necessary to provide networks

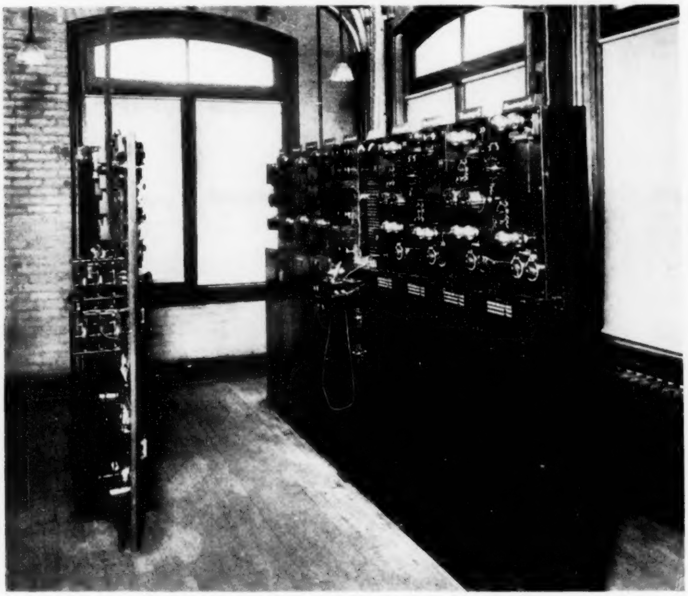


Fig. 1

to balance the lines over which the carrier system is operated. In the past both of these methods have been used but the tendency is now in the direction of eliminating balance entirely on account of its attendant maintenance difficulties and of providing for directional selectivity entirely by means of different frequencies in the two directions.

In order to show the variations in equipment arrangements which have been used in carrier systems, Figs. 1, 2 and 3 have been included. Fig. 1 shows one terminal of the original Pittsburgh-Baltimore carrier

telephone system. In this picture it will be noted that the apparatus is mounted on racks about 6 feet high and occupying about one square foot of floor space, which are lined up in rows as space permits. Fig. 2 shows a terminal of a later type of carrier telephone equipment which was installed between Harrisburg and Detroit. In this case the apparatus for a complete terminal (4 channels) is mounted on four relay bays as shown. The carrier telegraph equipment shown in Fig. 3 is a typical installation of the latest apparatus. Here rack construction is used although the individual panels are considerably larger than the older telephone equipment.

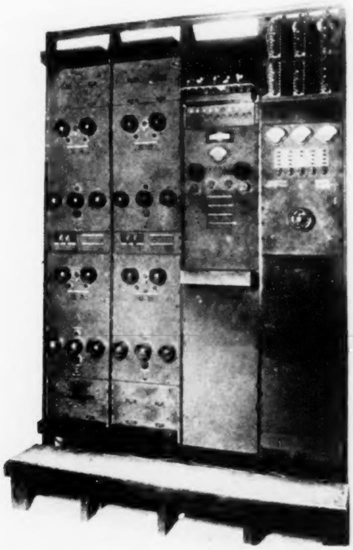


Fig. 2

Present Development

As pointed out by Mr. Vail in his original announcement of the successful development of the carrier equipment, carrier systems are economical only for the longer circuits in the plant. The cost of the terminal equipment is so great that short circuits cannot economically be provided by carrier apparatus. Repeaters, for amplifying the high frequency currents must also be installed at frequent intervals.

The permissible distance between these repeaters depends on the gauge of wire employed in the circuits on which the carrier circuits are superposed. For this reason the large gauge circuits of the Bell System have been equipped first with the result that practically all the existing carrier installations are installed on the 165 mil wires

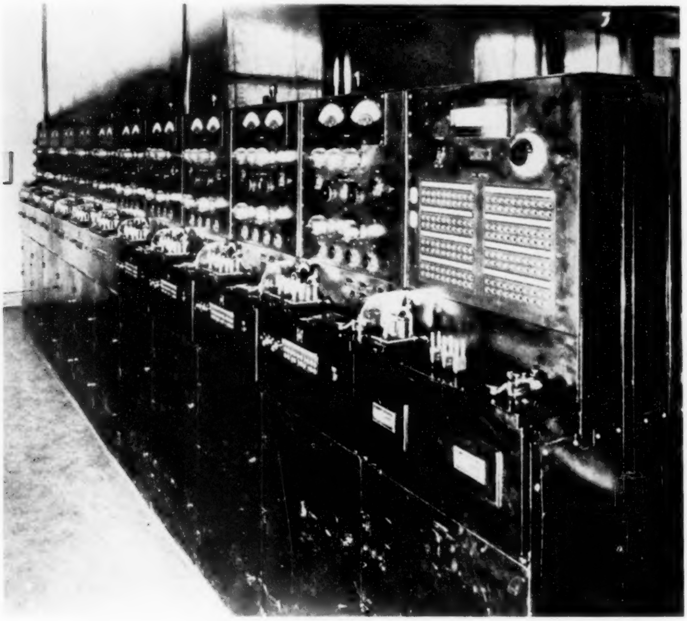


Fig. 3

which are the largest generally in use throughout the Bell plant. This wire is largely used on the important backbone routes of the country and it is on these that the existing carrier circuits are superposed. In looking over the map of Fig. 4, which shows all the existing carrier installations, this fact will be noted and also that the carrier systems in most cases provide circuits over 250 miles in length.

It is of interest to note that the application of carrier very completely covers the important cross country routes. In the west the circuits from Portland to Los Angeles are equipped, the transcontinental line from San Francisco east to Harrisburg, and the eastern coast route from Bangor to Atlanta with the exception of the all cable sections between Boston and Washington. As each line on the map represents several channels the number of circuits obtained by carrier

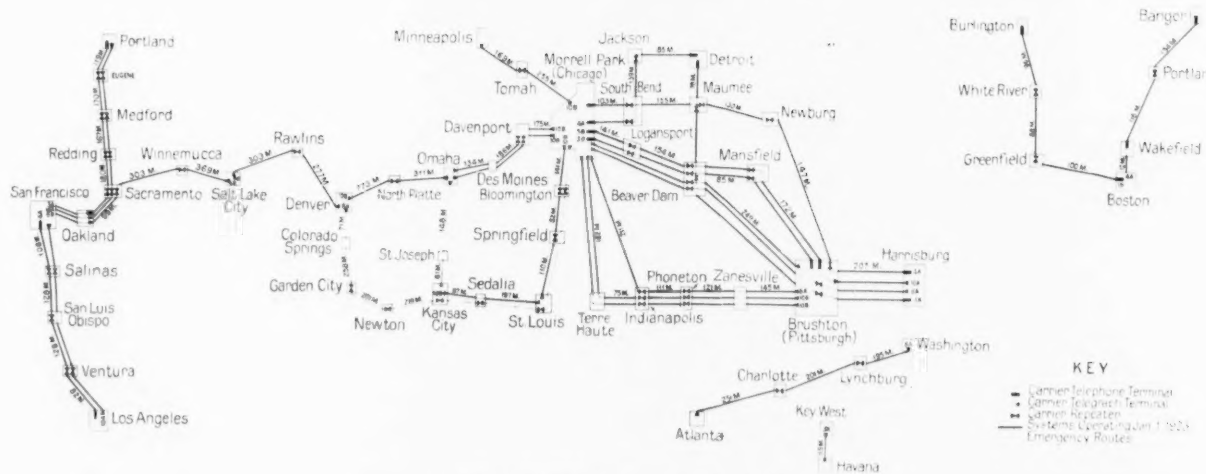


Fig. 4—Carrier System Installations

does not appear as large as is actually the case. In order to give a better idea of the extent of the carrier application the following table has been prepared which lists all the systems shown on the map and totals the channel miles obtained by each system. The telegraph channels if placed end to end would circle the globe 3 times and the telephone channels would extend somewhat more than half way round.

Carrier Telephone System	Channels	Miles	Channel Miles
Harrisburg-Chicago	4	742	2,968
Boston-Bangor	4	250	1,000
San Francisco-Los Angeles	4	446	1,784
Harrisburg-Detroit	3	605	1,815
Boston-Burlington	1	284	284
Oakland-Portland	3	735	2,205
Pittsburgh-Chicago	6	552	3,312
Chicago-Detroit	4	327	1,308
Total	29	3,941	14,676

Carrier Telegraph System	Channels	Miles	Channel Miles
Washington-Atlanta	8	647	5,176
Harrisburg-Chicago	18	749	13,482
Oakland-Portland	10	735	7,350
Chicago-Omaha	20	495	9,900
Chicago-Pittsburgh (Via Terre Haute)	20	634	12,680
Chicago-Pittsburgh (Via Indianapolis)	8	588	4,704
Key West-Havana	3	115	345
Chicago-Minneapolis	10	424	4,240
Chicago-St. Louis	10	333	3,330
St. Louis-Kansas City	10	294	2,940
Omaha-Denver	10	584	5,840
Denver-Salt Lake	8	580	4,640
Salt Lake-Oakland	6	771	4,626
San Francisco-Los Angeles	10	446	4,460
Total	151	7,395	83,713

Typical Cases—Telegraph

It will perhaps be of interest to consider several typical cases of carrier installations in order to see the economies involved in providing circuits by carrier rather than by other methods. Taking first the carrier telegraph systems as the considerations involved

there are usually very simple, we shall consider the Pittsburgh-Chicago section. There are at present three carrier telegraph systems actually in operation between Pittsburgh and Chicago. They provide a total of twenty-eight full duplex channels. These give service which could not be given otherwise as all the open-wire facilities between Pittsburgh and Chicago are completely equipped with direct current composited telegraph sets (to give all possible telegraph channels). The layout of the carrier telegraph systems between these points is shown in Fig. 5.

The above example is representative of the conditions under which carrier telegraph will be installed. In cases where open wire or cable facilities are available which can be composited with the ordinary direct current methods, the telegraph facilities can be obtained as a by-product most cheaply in this way. As soon as these facilities are all in use or an insufficient number of spare circuits remains, carrier telegraph can properly be used provided the returns from the special contract telegraph service are sufficient to meet the annual charges on the apparatus itself.

One of the carrier systems listed in the above tables is the Key West-Havana carrier system. The details of the telephone and telegraph channels obtained for the submarine cables were described in detail in the paper on the "Key West-Havana Submarine Telephone and Cable System" published in the journal of the A. I. E. E., dated March, 1922. On account of the considerable length of this cable and its high attenuation, the carrier equipment is special although resembling in principle the carrier telegraph apparatus used in our ordinary land installation. Without the carrier equipment it would have been possible to obtain only one telegraph channel on each of the three submarine cables by means of direct current composite sets. With the carrier apparatus it is possible to obtain 4 telegraph channels in addition to the single telephone channel.

Typical Cases—Telephone

The most important application at the present time of carrier telephone apparatus is probably between Pittsburgh and Chicago where the existing open wire leads are so congested that the additions of further circuits would require extensive construction work and possibly an entirely new pole line. An engineering study of this situation resulted in the drawing up of plans for an aerial toll cable which will largely replace the open wire. Work is already well advanced on the installation of this cable and it will be completed

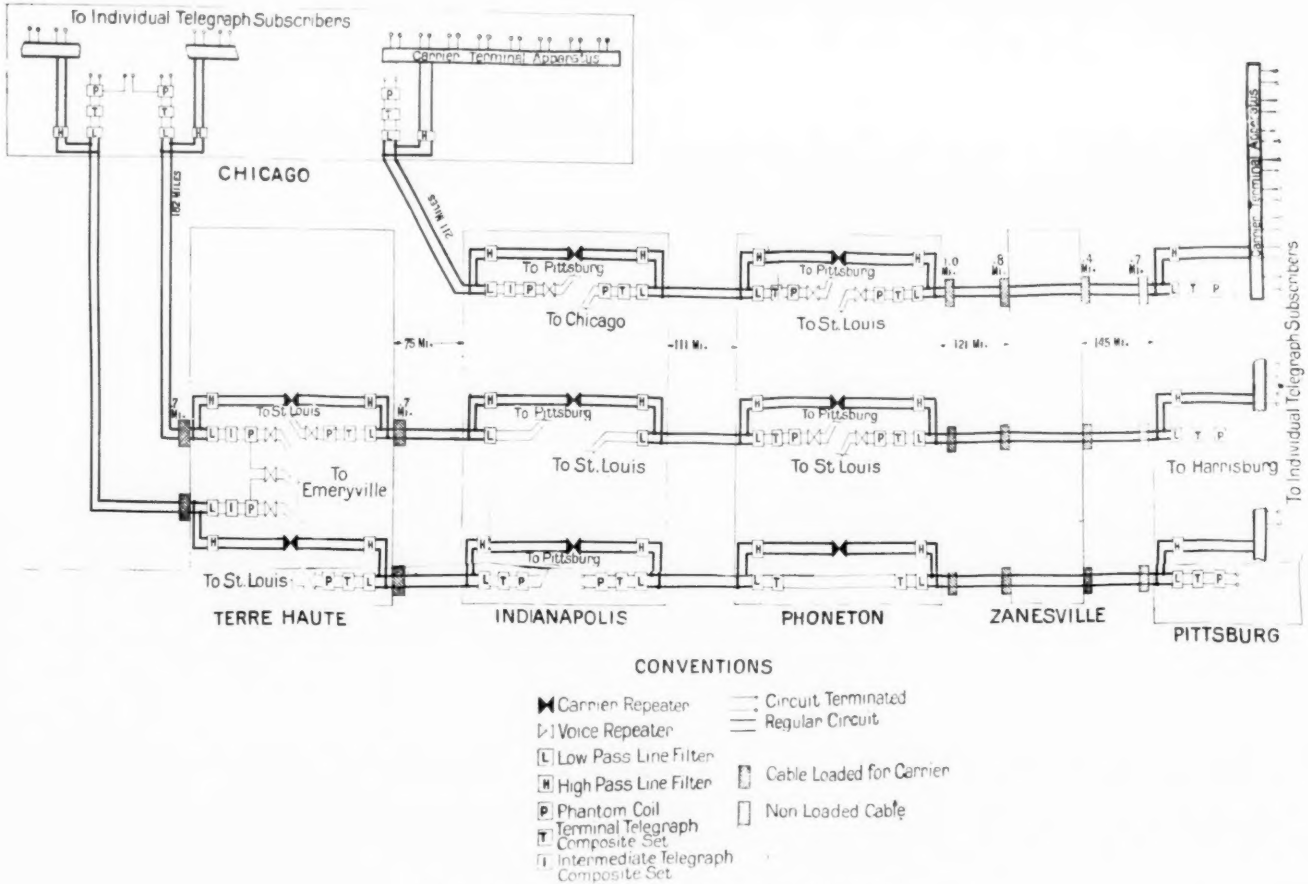


Fig. 5 Chicago-Pittsburgh Carrier Telegraph Systems

within the next few years but in the meantime, the use of carrier apparatus enables the traffic growth to be taken care of without stringing wire which, under the circumstances, would be very expensive.

There are, at present, operating between Harrisburg and Chicago, one 4-channel telephone system and between Pittsburgh and Chicago, two 3-channel systems, providing a total of 10 carrier telephone channels. These will be supplemented by at least two additional systems before the cable is completed. As soon as the cable is installed the carrier systems will probably be removed from service here and reinstalled in other locations.

In most cases the problem of providing additional circuits is not as difficult as in the section between Pittsburgh and Chicago. For this reason the relative economies of providing circuits by carrier and by the other methods must be more carefully considered. Even where no congestion exists, however, it will be found that where the circuits are long enough the carrier circuits will be cheaper than any other method of providing the facilities. The circuits to be provided must usually be several hundred miles in length before this is the case; also, since the cost of a carrier channel goes down as the number of channels installed at one time is increased it will usually be found that an installation of 3 channels will prove in for considerably shorter distances than would be necessary if a lesser number of channels are installed. In the practical case a complete system consisting of either three or four channels is usually installed at one time.

A typical case of a carrier telephone installation where the existing open wire lead is not already full but where the circuits required are long, is the Oakland-Portland system which in conjunction with a short cable between Oakland and San Francisco provides San Francisco-Portland circuits. The detailed layout is shown on Fig. 6. Here the cost study showed a considerable saving in annual charges in favor of the carrier although there was room for stringing open wire on the existing pole line. This system was put in service by the Pacific Telephone and Telegraph Company in 1921 and has since given very satisfactory service.

Another type of carrier installation is one installed to defer a proposed cable project. A long toll cable project involves the investment of such large sums of money that deferring the annual charge on the cable circuit for one year will frequently be sufficient to pay for and maintain a carrier system over the same period. Additional carrier systems may then be added to further defer the cable if this appears economical. The addition of a second system to a lead usually involves some considerable line expense for transposition work,

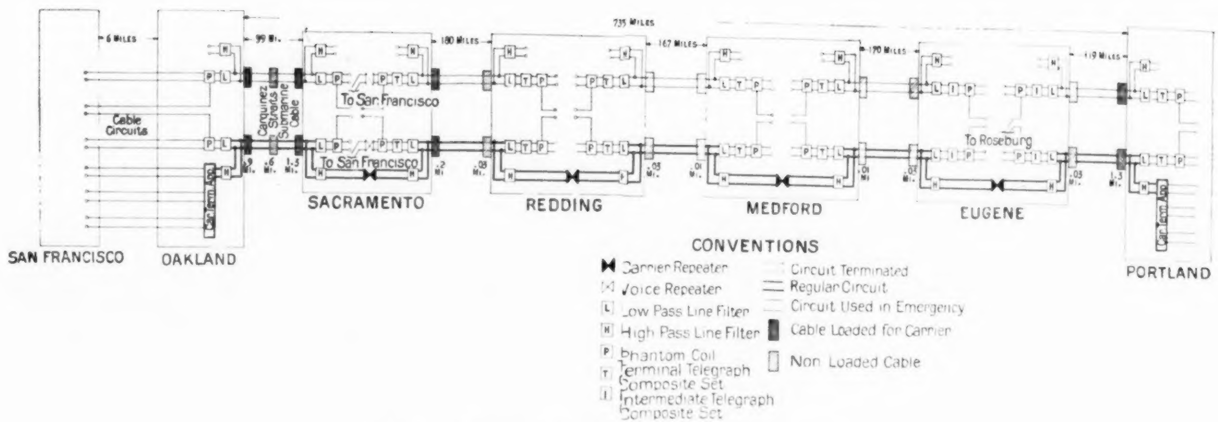


Fig. 6—Oakland-Portland Carrier Telephone System

however, and this may prove out the further addition of carrier. Even if it appears economical to install additional carrier systems, a point will soon be reached where the cumulative annual charges of the carrier systems will exceed that of the cable. The first few systems prove in over the cable because the carrier provides only for the immediate circuit requirements while the cable must take care of growth and therefore includes many idle facilities when first installed. Where carrier is used to provide facilities in place of a toll cable it should always be considered an intermediate and temporary step between open wire and cable plant.

The use of carrier as outlined above may effect further economies after the apparatus has been removed as the equipment may be reused at some other point to advantage. A typical example of the use of carrier apparatus to defer a cable is the Boston-Bangor carrier system which was put in to defer the installation of the first section of the Boston-Portland cable. The layout of this system is given in Fig. 7. It will be seen that this system is fairly short but the first

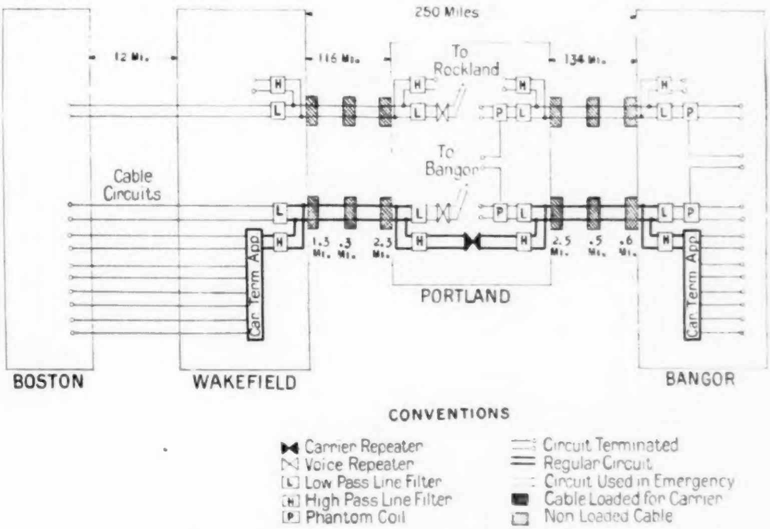


Fig. 7—Boston-Bangor Carrier Telephone System

year's annual charge on the first 50 miles of the Boston-Portland cable would have been sufficient to pay the entire first cost of this carrier project. It is possible that further carrier may be installed on this route before the cable is finally installed. The present system

has deferred the cable somewhat longer than was originally expected as the growth of traffic has not been quite as rapid as was expected at the time of the war emergency.

Another example of line congestion enabling the carrier to be proved in on somewhat shorter than the ordinary economical length is in case a considerable amount of line reconstruction is involved if open-wire circuits are added to an existing lead. A case of this kind was the Boston-Burlington system where a very considerable amount of line reconstruction work would have been involved if an effort had been made to add a phantom group to the existing lead. The use of the carrier system on the existing 104 mil circuits enabled this work to be eliminated from consideration and it is possible that the work will not need to be done until this section of the line is relieved by cable or other means.

There are many cases in which the use of carrier can be considered a stop-gap to take care of the transient period between open wire and cable facilities. This has been true in the case of the former Baltimore-Pittsburgh system where the original apparatus has been removed from service as cable facilities are now available between these points via Philadelphia, Reading and Harrisburgh. This does not mean that the equipment is no longer of value, since it usually can be used again on some other location. Even the experimental panels which were used in the Pittsburgh-Baltimore system will probably be reinstalled within the next year. It is now thought that this apparatus will be used between Chicago and Minneapolis in connection with some additional panels to provide for new telephone circuits there.

Expected Development

Looking forward for the next ten years, it is expected that carrier telephone facilities will be installed at the rate of about 5,000 to 10,000 channel miles and telegraph facilities at the rate of from 20,000 to 30,000 channel miles per year. In the meantime development work may produce cheaper systems which will prove in on shorter circuits, thereby extending the field of use so that the rate of application may possibly be doubled or trebled. Even now the number of channel miles in service constitutes an important part of the total facilities of the Bell System and present a very interesting picture of rapid growth when compared with the beginning between Baltimore and Pittsburgh in 1918.

Machine Switching Telephone System for Large Metropolitan Areas

By E. B. CRAFT, L. F. MOREHOUSE and H. P. CHARLESWORTH

SYNOPSIS: From the earliest forms of telephone switchboards to the modern types, the development of the switchboard has been marked by the increasing use of automatic methods to supplement the manual operation wherever this would result in better service to the public or more efficient operation.

In addition to all that has been done in developing and introducing automatic operations with manual switchboards, it has been found desirable and practicable to go further in the direction of introducing automatic operation in the telephone plant and a machine switching system has been developed in which the bulk of the connections are established without the aid of an operator.

The complexity of a large metropolitan area and the exacting requirements which a machine switching system must meet are outlined briefly, and the system which has been developed to meet these requirements is described.

The application of the system to a typical large metropolitan area and the means provided for permitting its gradual introduction into the existing plant are discussed.

IT is the purpose of this paper to outline briefly certain important developments in connection with machine switching telephone systems and to discuss the application of the results of these developments to the problem of providing telephone service in large metropolitan areas.

The telephone was invented in 1876. Almost immediately thereafter it was recognized that, for it to attain its greatest field of usefulness, switchboards and switching centers would have to be established for effecting interconnection between subscriber's lines.

Professor Bell's vision of the future was given in a statement to prospective investors. He said:

> "It is conceivable that cables of telephone wires could be laid underground, or suspended overhead, communicating by branch wires with private dwellings, country houses, shops, manufactories, etc., etc.—uniting them through the main cable with a central office where the wires could be connected as desired, establishing direct communication between any two places in the city. Such a plan as this, though impracticable at the present moment, will, I firmly believe, be the outcome of the introduction of the telephone to the public. Not only so, but I believe in the future wires will unite the head offices in different cities, and a

Presented at the Midwinter Convention of the A. I. E. E., New York, N. Y., February 14-17, 1923. Published in the Journal of A. I. E. E. April, 1923.

man in one part of the country may communicate by word of mouth with another in a distant part.

"Believing, as I do, that such a scheme will be the ultimate result of the telephone to the public, I will impress upon you all the advisability of keeping this end in view, that all present arrangements of the telephone may be eventually realized in this grand system."

Early Developments

The only apparatus available at that time for this purpose was that employed in telegraph, messenger, fire and burglar alarm services. Some of this apparatus, such as wire, insulators, batteries, annunciators, etc., was found to be useful in the new art; other apparatus had

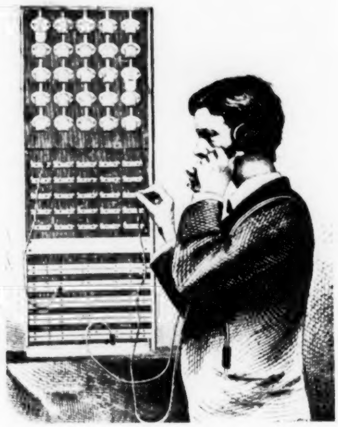


Fig. 1—Early Type Switchboard

to be developed. The switchboards of that day employed this apparatus. They were small in size, and could accommodate only a limited number of lines.

It soon became evident that the requirements of the telephone exchange service demanded signaling and switching equipment different from that employed in any of the other branches of the electrical industry, and it became necessary to create an entirely new art, involving many branches of science, before commercial telephone service could be given on an adequate scale. The switchboards grew from small boards, capable of handling a few lines, as

shown in Fig. 1, to the very complex arrangements providing signaling, switching, and transmission facilities for as many as ten thousand lines in a single board, of the type shown in Fig. 2.

As the subscribers increased in number it was found that beyond a certain point it was no longer practicable or economical to have all of the subscribers' lines brought to one center. It was therefore, necessary to have several centers, the number depending upon many factors, the most important of which are the size and telephone needs of the community.

Fig. 2—Modern Type Common Battery Switchboard

The consequence of all this is that in large metropolitan areas the number of centers is large, and the trunking system complex, as each center must be provided either directly or indirectly with trunks to every other center.

As an illustration, take the New York Metropolitan area, shown in Fig. 3, where the telephone plant is of the greatest intricacy because of the very large number of subscribers served. There are at the present time 158 central office switchboards, many of them having

equipment for 10,000 lines. These offices and the associated plant provide for intercommunication between 1,400,000 telephones, and approximately two trillion possible connections. It is estimated that by the year 1940 there will be 300 central office switchboards within the New York Metropolitan area, serving some 3,300,000 telephones—or nearly two and a half times the present number.

Fig. 3—Map Showing Location of Central Offices in New York Metropolitan Area

Manual Switchboards

The system most commonly employed today for connecting subscribers' lines together is the so-called "manual" system; that is, a system in which operators are employed to make the actual connections between subscribers' lines, although so many of the functions are performed automatically that, except in name, it is to a large degree automatic.

It is a long step from the early switchboards to the modern common battery multiple manual switchboards. The history of the development of switchboard equipment and apparatus shows that enormous progress has been made in this art in a comparatively few years. As the telephone subscribers have grown in number and as the amount and complexity of the traffic have increased, it has been only by the most intensive development that it has been possible to keep ahead of the demand for telephone service, and that telephone engineers have been able to get the speed, efficiency and accuracy that are obtained

today in so-called manual operation. It is worthy of note in this connection that the attainment of these ends was made possible by the extensive introduction of automatic features.

A very brief description at this point of the type of manual switchboard more commonly employed will be helpful.

In this switchboard the subscriber's line terminates at the central office in so-called "jacks." Associated with each line is a lamp, individual to it, which automatically lights when the subscriber removes his receiver from the hook. This serves as a signal to the operator that a connection is desired.

The operator answers this call by inserting one end of a cord in the jack associated with the calling subscriber's line, operates a listening key which connects her telephone set to the subscriber's line, and asks for the number desired. When this is obtained the operator completes the connection by inserting the other end of the cord in the jack of the desired subscriber's line, and the subscriber's bell is rung. Suitable lamp signals are provided so that the operator may know when the called subscriber answers, when either subscriber desires further attention, or when either or both of them have finished talking and have hung up their receivers.

If the subscriber desired is connected to a distant office, the operator receiving the call would, instead of plugging directly into the subscriber's line, directly connect the subscriber's line to a trunk terminating in the desired office, where the connection would be completed by a second operator, known as the "*B*" operator, as shown in Fig. 4.

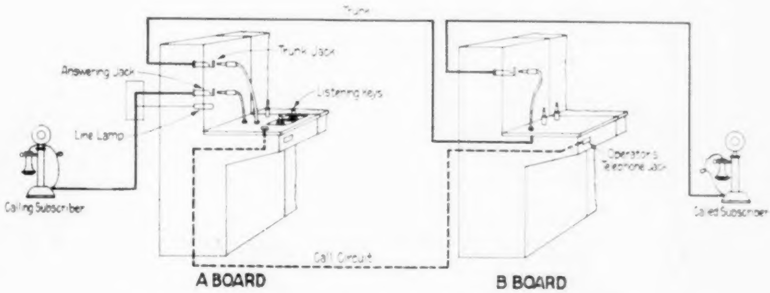


Fig. 4—Diagram Showing Manual Interoffice Connection

Such communication between the two operators as is necessary to establish this connection takes place over a special pair of wires known as a "call circuit."

The method, of which the above is a bare outline, is that used in completing ordinary connections. Different arrangements and different operating methods have to be provided for handling short haul toll calls, long distance calls, calls from coin boxes, and calls of many other kinds.

In the simplest types of manual systems, the subscriber, in order to signal the central office, turns a crank thus operating a magneto generator. This throws a drop in front of an operator at the central office. In the switchboards developed to meet the needs of the larger areas, electric lamps are substituted for the drop, and relays automatically controlled by the subscriber bring them into play at the proper time. Electric lamps which serve as visual signals to the operator to indicate the status of the connection are also associated with the cords that the operator uses for connecting subscribers together. The operation of these lamps is automatic and is under the control of the switchhook at the subscriber's station.

Many other arrangements of an automatic character have been developed and are used as occasion requires—not merely because they are automatic in character but only when it has been established that they make for better service to the public or for efficiency and economy of operation, or both. Among these may be mentioned automatic ringing, automatic listening, and many forms of automatic signaling. Many of these arrangements are highly ingenious and contribute greatly to the efficiency and economy of operation. Thus, the trend of switchboard development has been more and more in the direction of automatic operation and automatic methods.

In addition to all that has been done in developing and introducing automatic operation with so-called manual switchboards, it has been felt for a long time that in large and complex telephone areas, such for example as New York City, the time would ultimately come when it would be desirable to go further in the direction of introducing automatic operation in the telephone system. This whole matter has been the subject of much thought on the part of engineers of the Bell System and, as a result, there has been developed and recently put into operation a system in which the work of establishing most of the local connections is done entirely by machinery.

The introduction of this system will eventually make a considerable reduction in the telephone company's requirements for operators which are becoming more difficult to fulfill year by year. Operators will be required, however, to handle toll and many special classes of local calls and for this reason, together with the constant growth of the business and the considerable period of time that will be re-

quired to introduce the new system completely, we can expect little or no reduction in the present operating forces for some time to come, and no operator will find herself out of employment on account of the introduction of the machine switching system.

Machine Switching

It is the purpose of this paper to describe this system sufficiently in detail to give a general picture of it, but because of the limitation as to space no attempt will be made to go into the intricacies of circuits and apparatus, which doubtless would be of interest only to the telephone engineering specialists.

Among other requirements, the following must receive special consideration in the design of a machine switching system.

The functions to be performed by the telephone subscriber in getting a connection must be simple and easily understood.

It must work efficiently and with accuracy and speed, and, of course, must be capable of handling the various types of calls that the subscriber wishes to make.

The system must not require modifications in the existing rate structure, otherwise than desirable. If the rate structure calls for message register operation, coin boxes, etc., means must be provided for automatically operating the register and collecting the coins on such calls, and for preventing a charge on calls not answered, calls for free lines, busy lines, etc.

The system should employ, as nearly as practicable, the conventional numbering scheme.

It should work with the existing telephone network, so that its introduction does not require wholesale number changes and extensive rearrangement or the abandonment of existing switchboards or other plant. Its introduction must, of necessity, be on a gradual basis.

It must be sufficiently flexible in design to care for growth and such changing traffic conditions as occur from time to time.

In large telephone areas, such as the New York Metropolitan area, there is a great variety of calls to be handled and many different classes of service furnished the public, such as message rate, flat rate, official, coin box, non-attended pay station, attended pay station, special services such as information, etc. Not only individual lines but party lines, and private branch exchanges must be cared for, and provision must be made for thousands of toll messages which must be recorded, supervised and timed.

A call originating in a machine switching office in New York City may have as its destination any one of a great number of points. It

may be for another subscriber in the same office or for one in another nearby machine switching or manual office; it may be for one of a large number of suburban toll points, or it may be to some point in a distant city.

The machine switching system, which is the subject of this paper, meets these requirements. After long-continued laboratory experiments, supplemented by field trials, power-driven apparatus of the panel type has been found to be the most suitable, and is now in successful operation in New York City and in other large cities in the country.

Fig. 5—Desk Stand Equipped with Dial

General Plan of Operation

At the expense of some repetition it seems desirable in order to give as clear an understanding as possible as to the operation of the system, to first give a brief outline of how the call is handled and a description of the more important elements of the equipment, before going into a detailed description of the operation of the system.

The subscriber's station is equipped with the usual form of telephone instrument and, in addition, with a calling device known as a "dial," mounted at the base of the desk stand, as shown in Fig. 5. This dial has ten finger holes, bearing letters and figures, as shown in Fig. 6.

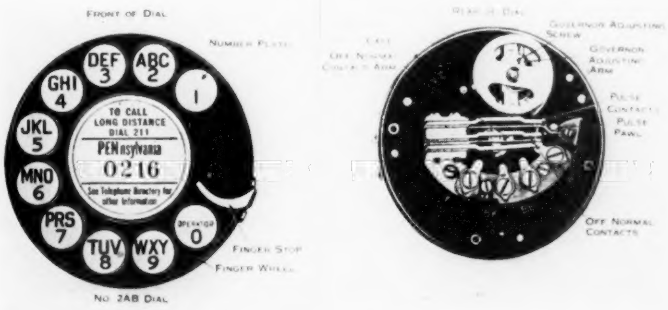


Fig. 6—Subscriber's Dial

In making a call the subscriber will, of course, first refer to the telephone directory. He will find in the directory a listing that is only slightly different from that to which he is accustomed. Typical samples of this new form of telephone listing for New York City are shown in Fig. 7. As will be noted, these conform to the present

Argent Co, 1400 Bway..............	**GRE** eley	5513
Argentina Brazil & Chile Shipping Co 70 Wall.	**HAN** over	0307
Argentine Genl Consulate, 17 Batry pl..	**REC** tor	6948
Argentine Impt & Expt Corp, Prod Ex...	**BRO** ad	1768
Argentine Mercantile Corp, 42 Bway.....	**BRO** ad	5066
Argentine Naval Commission, 2 W 67..	**COL** mbus	5623
Argentine Quebracho Co, 80 Maiden la....	**JOH** n	1652
Argentine Railway Co, 25 Broad.......	**BRO** ad	1383
Argentine Trading Co, 1164 Bway......	**MAD** Sq	1871
Argeres Bros, Restrnt, 86 6th av......	**SPR** ing	5337
Argero A, Grocer, 119 9th av........	**CHE** lsea	6255
Arghis A, Tobacco, 74 Wall.........	**HAN** over	6311
Argiropie Theodore, Jwlr, 406 8th av..	**FAR** ragut	9772
Argo Packing Corpn, 705 Greenwich...	**FAR** ragut	4505
Argon Dress Co, 24 E 12..........	**STU** yvsnt	2011
Argonaut Supply Corp, 50 Union sq..	**STU** yvsnt	7476
Argonne Steamship Co, 17 Battery pl...	**REC** tor	2493
Argos Ad-Art Co, 1133 Bway........	**FAR** ragut	5986
Argosy The (A Pub), 280 Bway......	**WOR** th	8800

Fig. 7—Typical Examples of New Form of Listing Telephone Numbers

manual listings, except that the first three letters of the office name are set out prominently. This numbering system will be discussed later in this paper.

Having secured the desired telephone number from the directory, which we will assume is "ACAdemy 1234," the subscriber will first remove his receiver from the hook and will hear the so-called "dial tone," which indicates that the apparatus is ready to receive the call. He will then insert his finger in the hole over the letter *A*, rotate the dial until the finger comes in contact with the metal stop shown in the picture, then release the dial, which will automatically return to normal. He will repeat this operation for the letters *C* and *A*, and in turn for the four numerals, *1, 2, 3, 4*.

This operation of dialing on the part of the subscriber is exactly the same, whether the telephone number he desires is in a manual or a machine switching office. Similarly, the method employed by a subscriber who is connected to a manual office in getting a subscriber connected to a machine switching office, is the same as though the desired subscriber were connected to another manual office.

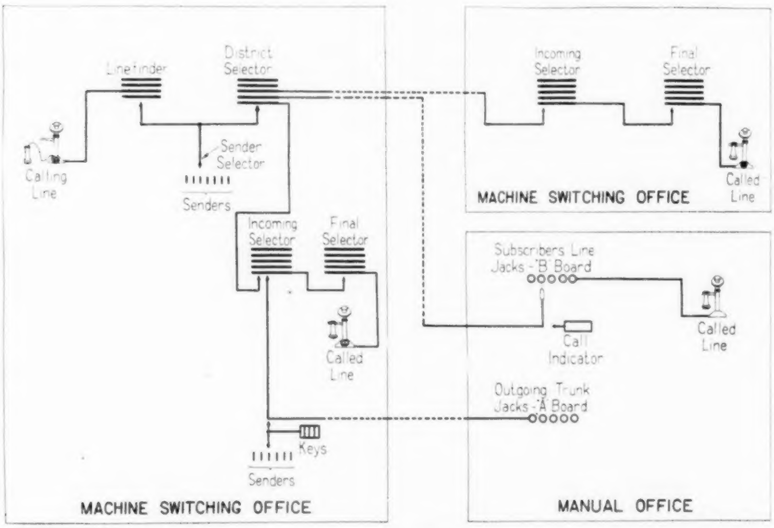


Fig. 8—Diagram Showing Connections from Machine Switching to Machine Switching, Machine Switching to Manual and Manual to Manual Switching

The progress of a call originating in a machine switching office is briefly as follows:

As will be seen from Fig. 8, the line of the calling subscriber, whom we will assume to be a subscriber in the Academy office, appears in a so-called "line finder" frame. When the subscriber's receiver is

removed from the switchhook preparatory to dialing, the line is selected by a "line finder" and connected to an idle "sender" by means of a "sender-selector."

Upon completion of these operations which take but a fraction of a second, the dial tone is sent to the calling subscriber as previously mentioned. When the subscriber dials, electrical impulses on a decimal basis are transmitted to the sender which receives and registers them, translating them in turn to the proper basis for the control of the selectors which are not operated on a decimal basis. The sender automatically causes the particular "district selector" which is permanently associated with the line finder originally used, to select a trunk to the office desired.

Assuming that the call is for a subscriber in the same office, Academy, the trunk chosen will terminate at an "incoming selector" frame and the sender above referred to will cause the call to be routed through the incoming selector to a final selector, and thence to the particular line desired. When the connection is thus completed, audible signals will be sent back to the calling subscriber to indicate that the station is being rung or that the line is busy.

If the call had been for a subscriber in another machine switching office, namely, Pennsylvania, the call would be routed from the district selector to the office desired, either directly or through an "office selector" in case the total number of trunks to all offices is too large to be placed on the district selector multiple. These trunks terminate on incoming selectors at the Pennsylvania office which select the subscriber's line through final selectors, as described above.

If the call is for a subscriber connected to, say, the Worth Office, which is a manual office, the call would be routed from a district selector directly or through an office selector to the "B" board in the Worth Office, where the number desired appears in front of the operator at a "call indicator position" in the form of visible numbers on the keyshelf. The operator is advised of the trunk to which the call is connected by suitable signals, and the call is completed by plugging this trunk into the desired subscriber's line.

Calls originating in a manual office and intended for a machine switching office reach the machine switching office over trunks from the "A" operators in the manual office. At the machine switching end these trunks terminate in incoming selectors, which have access to the final selectors on which the subscriber's lines are located. The selectors are under the control of a special group of senders, and operators are provided with suitable keys for setting up in these senders the number of the desired subscriber. These operators at

the machine switching office receive the information as to the desired number from "A" operators in the distant manual office, exactly as is done in the case of manual operation.

The introduction of machine switching equipment does not require radical changes in private branch exchanges. The private branch exchange is provided with dials, and calls to the central office are dialed by the private branch exchange attendant or by the extension user in the same way that the ordinary subscriber dials. No change in the private branch exchange is required for handling incoming calls. An idle trunk in the private branch exchange group is selected by the mechanism in the machine switching office, in much the same way as an individual subscriber's line is selected.

Numbering System

One of the unique advantages of the plan developed for designating telephone numbers, to which reference has already been made, is that it does not necessitate the abandonment of the existing manual listings. It requires no change except that the first three letters of the office name are set out more prominently. Simple as this change in the form of listing appears, until it was developed by the Bell System no satisfactory method of designating telephone numbers for machine switching offices in large cities was known.

Many plans had been proposed, to all of which there were serious objections. Some of them required changing the whole system of manual designation, others the use of combinations difficult for the subscriber to use. In small cities a numbering plan employing only digits is sometimes practicable, but in such a large area as we are considering, such a plan would involve seven digits. The subscriber's number would take the form of say 786-3549. Such numbers would be difficult for operators to use and for the subscribers to carry in mind and would require that every subscriber's number in the entire area be changed before the first machine switching office could be cut into service. With the new system, the subscriber's number and office in general remains as before. It is necessary to change only a few conflicting office names in order to make them fit into the system.

Description of the Equipment

A detailed description of each unit employed in this system would be impracticable, in this connection, but a brief description of the more important ones will be of interest.

Sender. The use of the sender makes practicable the introduction of machine switching in large metropolitan areas where, of necessity,

the service conditions are extremely complex. It is, in effect, the brains of the system, dealing with the subscriber and controlling the selection until the destination is reached, as an operator deals with the subscriber and controls the selection in a manual system. The number dialed conveys the same information to the sender in a machine switching system as the number spoken by the subscriber does to an operator in a manual system.

The sender is an arrangement of relays, sequence switches, and selectors, so worked out as to perform the following more important functions:

1. It receives a succession of electrical impulses from the subscriber's dial which are on a decimal basis, stores them and translates them to a non-decimal basis, corresponding to the particular group of lines and trunks that is involved in the path of the call.

2. It controls selecting mechanisms which build up the connection to the called party in such a manner that each mechanism is given the exact time required to perform its functions without any waste of time, independently of the rate received from the dial.

3. It makes the central office designations entirely independent of the arrangement of the trunk groups on the selector frames. This is a very important matter, inasmuch as it allows the selectors to be used to full efficiency. It provides the desired flexibility for growth and permits any desirable rearrangement of the trunks on the selector frames that the telephone company may find desirable at any time.

4. The sender is capable of distinguishing the class of office at which the connection terminates. That is, if the call is to terminate at a mechanical office, the sender will arrange to govern the selection accordingly. If the call is to terminate at a manual office, the sender recognizes this and arranges to send out impulses to the call indicator equipment in the manual office.

5. For the completion of certain calls, traffic conditions require the introduction of tandem centers as discussed later. The sender recognizes calls to be routed via tandem centers and arranges to handle these correctly. The tandem center may be manual or it may be mechanical, and the control must be determined accordingly.

6. Certain senders are arranged to serve lines supplied with coin boxes. These senders are arranged to make a test to determine whether a coin has been deposited and do not allow the connection to be cut through so that conversation can take place until the coin is deposited. If the subscriber does not deposit the coin, after a reasonable time has elapsed the sender connects an operator to the subscriber, and this operator notifies the subscriber of his omission. After the

coin has been deposited, the sender allows the called subscriber to be rung and permits the conversation. In case the called subscriber is busy or does not answer, or if the call is to a free line, the sender returns the coin to the calling party. If the called party answers, the sender causes the coin to be collected.

The sender makes a test of the calling line after the subscriber has completed dialing, to insure the deposit of the coin, and recognizes whether a coin has actually been deposited or whether some abnormal condition exists, in which case the call will be routed to an operator who causes an investigation to be made.

7. In large areas, such as the New York Metropolitan area, there are distant points, connection to which requires toll charges. In such cases the subscriber is instructed to dial a special operator who will ascertain his wishes, complete the call, and make the proper charge. Should a subscriber attempt to dial outside of his own local service area, his call will automatically be routed to an operator.

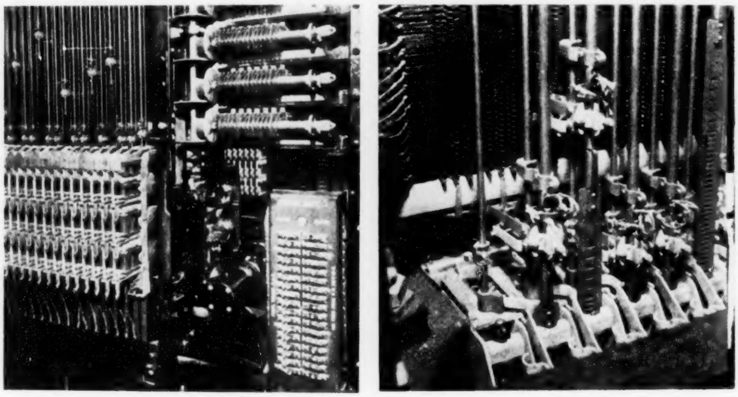


Fig. 9—Views of the Selector

Panel Type Selecting Mechanism. An important mechanism of a machine switching system is the selector and its associated multiple bank. It is a device by means of which trunks or lines are connected together as required. It performs the same function as the switchboard cord and plug which in a manual exchange can be plugged by the operator into any one of a number of jacks which are the terminals of trunks or lines.

Fig. 9. shows the mechanical elements of the selectors. The movable member corresponds to the cord and plug of the manual system

and the fixed terminals or multiple, to which the movable member can make connection, corresponds to the jacks of the manual system.

Fig. 10 shows the fixed terminals or multiple to which the selectors connect. This multiple consists of flat punchings about 3½ feet long and 1 inch wide overall. Each of these strips has lugs on each side with which the selectors can make contact. In this particular panel, three hunderd of these strips are piled one above the other, separated by insulation, and securely bolted together, forming a panel about 15 inches high. This panel provides a multiple consisting of "tip," "ring" and "sleeve" connection for one hundred lines appearing sixty times; that is, thirty on each side. The insulating material consists of special impregnated paper and is of such a nature that,

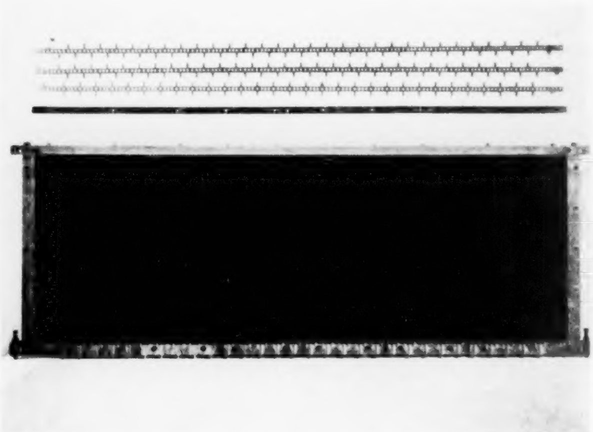


Fig. 10—Selector Multiple Bank

after the panel is assembled and baked, it becomes inert and is not adversely affected by any conditions met with in a central office. It is this panel which has given the name to the system.

The selector (see Figs. 9 and 13) consists of a metal tube supported in bearings allowing vertical motion and carrying five sets of brushes. Each one of the five sets of brushes is arranged to make connections to the tip, ring and sleeve terminals of the panel banks before which it normally stands, and the tip, ring and sleeve contact members of all five of these brushes are multipled together. They are normally free from contact with the terminals, but any set may be tripped me-

chanically, so that that set will contact successively over terminals as the selector rises.

A friction clutch is provided at the base of each selector, so arranged that the selector can be raised or lowered by power supplied by a constantly rotating small motor, common to 60 selectors. A magnet is also provided for tripping, by means of a rotating rod, any one of

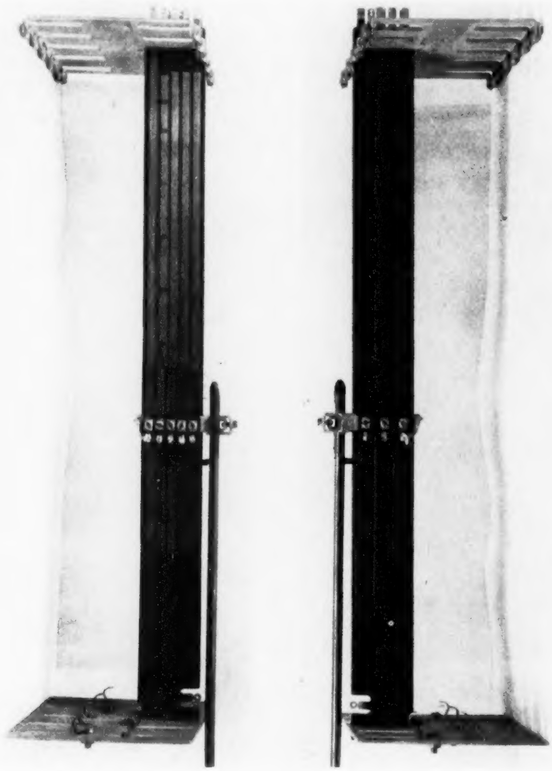


Fig. 11—Commutator for Controlling Vertical Movement of Selecting Mechanism

the five sets of brushes into mechanical engagement with the terminals. In choosing a trunk or line, that one of the five sets of brushes which has access to the panel in which the desired trunk or line happens to be, is tripped so that it makes contact with the bank terminals before it. The selector then moves upward, under the proper control,

until the tripped brush engages the desired line or trunk. The selector is then held in this position by a pawl associated with the clutch. When the connection is to be taken down, the pawl is withdrawn, and the selector is carried back by means of the power drive controlled by the clutch. When the selector reaches its normal position the tripped brush is reset.

Selectors used for different purposes are arranged to move their brushes upward at different speeds. The speed most commonly employed moves the brushes over the terminals at the rate of 60

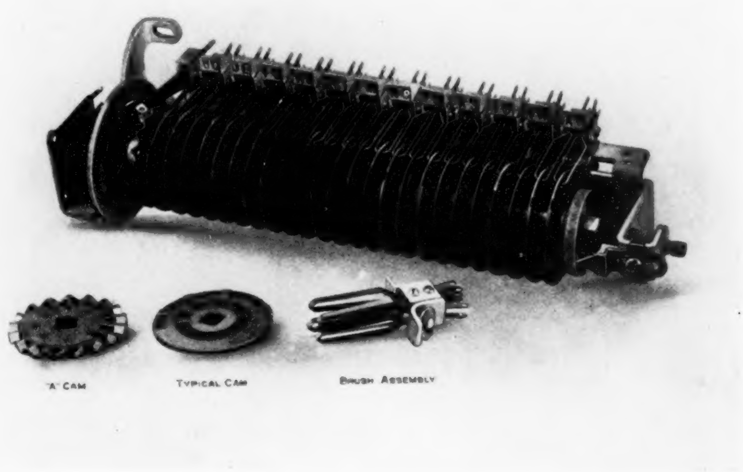


Fig. 12—Sequence Switch Assembly

trunks per second. At the top of the frame, just above the fifth bank, are located commutators as shown in Fig. 11, one for each selector. The multiple wiring of the brushes on the selector leads to other brushes which move over strips on these commutators, and thereby completes the connection from the movable selector to the rest of the circuit, thus avoiding flexible wire connections with their attendant troubles. This commutator, also, performs the more important service of controlling the travel of the selector. Brushes moving over conducting segments separated by insulation produce

impulses which, when sent back to the sender, indicate to it the exact position of the selecting mechanism.

Sequence Switch. Another device of great importance is the "sequence switch," shown in detail in Fig. 12. It is operated through

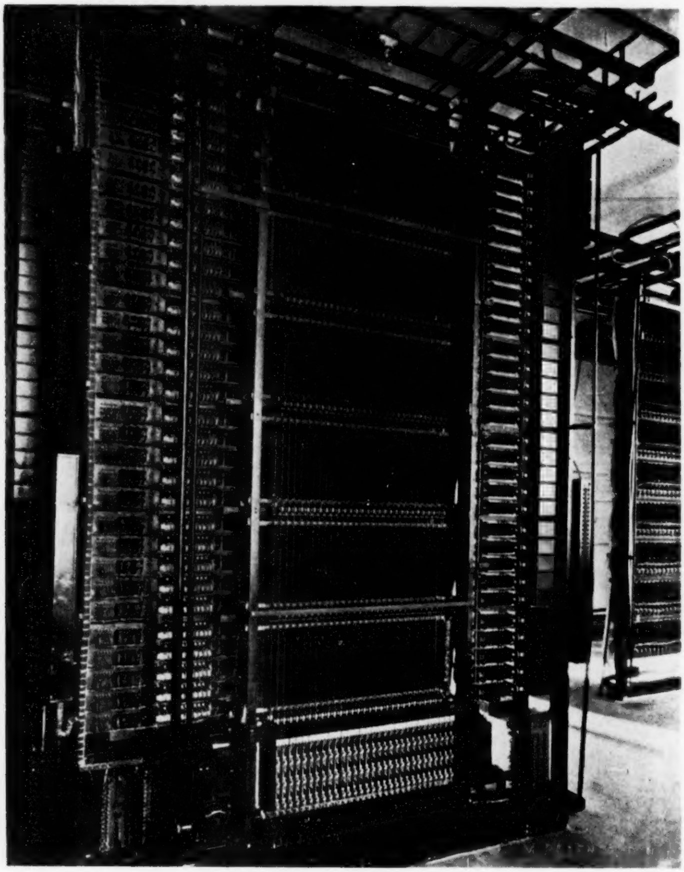


Fig. 13—Selector Frame Completely Equipped

an electromagnetic clutch from the same motor that drives the selectors.

The sequence switch may be described as a circuit controller or device whose function is to establish in a definite sequence such circuit

conditions as are required in the operation of the system. It is made up of circular disks called cams mounted rigidly on a shaft. The plates of the cams are cut so that brushes come in contact with the plates only when the circuit is to be closed. The sequence switch can be stopped at any one of eighteen different positions as required, by the simple opening of the electromagnetic clutch.

There are many of these sequence switches used in this system, and the arrangement of cutting the cams varies, depending upon the particular circuit combinations which it is desired to establish.

Selector Frames. Fig. 13 shows thirty selectors with all of the associated mechanism mounted upon one side of a frame ready for operation in an exchange. Both sides of the frame are alike. Five panels of 100 lines each are mounted in this frame, one above the other giving a total capacity of 500 trunks or lines. Thirty selectors, each capable of making connection with any one of the 500 trunks or lines, are placed adjacent to each other on each side of these panels; the entire frame thereby having a capacity of sixty selectors, each of which has access to 500 trunks or lines.

Immediately to the right of the selectors are the sequence switches and, under protective covers, such relays as are used in connection with the selectors upon the frame shown.

Selecting apparatus of this general type, but differing in details of design, is used during the different stages of the call as line finders, district selectors, incoming selectors and final selectors, reference to which has been made before. Fig. 14 shows a section of a machine switching office with some of the typical frames.

The use of apparatus of the substantial construction just described is made possible only through the use of the sender which receives impulses from the subscriber at the rate they are dialed and receives impulses from the selecting mechanism at the rate it is traveling. This obviates the necessity for restrictions in the design of either the dialing circuit or the selecting circuit, such as would be necessary if they were tied together.

Power Supply Arrangements. Since most of the operations normally required in handling a call in a machine switching office are carried out mechanically, it is evident that a considerably larger amount of power is required than with the manual system. Selectors and sequence switches are propelled mechanically by rotating shafts driven continuously by small motors mounted on each frame.

The use of small motors on each frame gives a flexible and reliable source of power particularly since the motors now being used are of the special "duplex" type developed for the purpose. They consist

of two motor elements in one frame, one element being normally driven from the commercial power service and the other being driven by the telephone reserve storage battery to which it is automatically

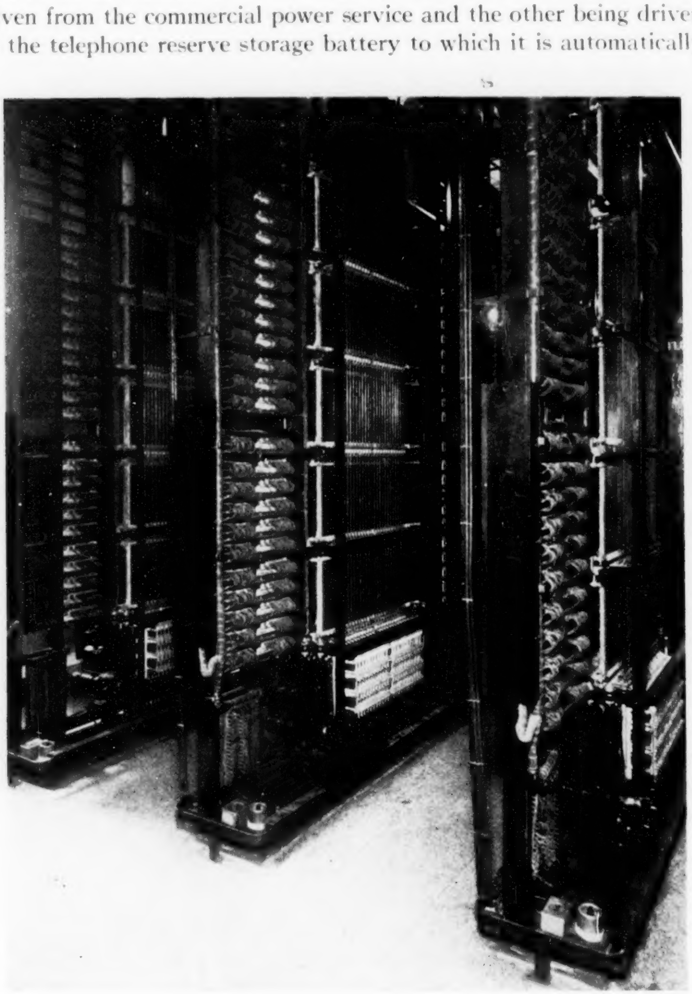


Fig. 14—Group of Typical Selector Frames

connected by a relay inside the motor when the regular power fails. A power failure, therefore, causes no interruption to the drive. The selectors are arranged so that not more than half in any one group are

driven by the same motor which insures continuous service in case of motor failure.

The main power requirement is for direct current at about 24 and 48 volts which is furnished from motor generator sets (Fig. 15) of special construction to reduce noise, converting the commercial alternating or direct power current into current which is regulated as to voltage and is free from variations which would cause noise in the telephone circuits.

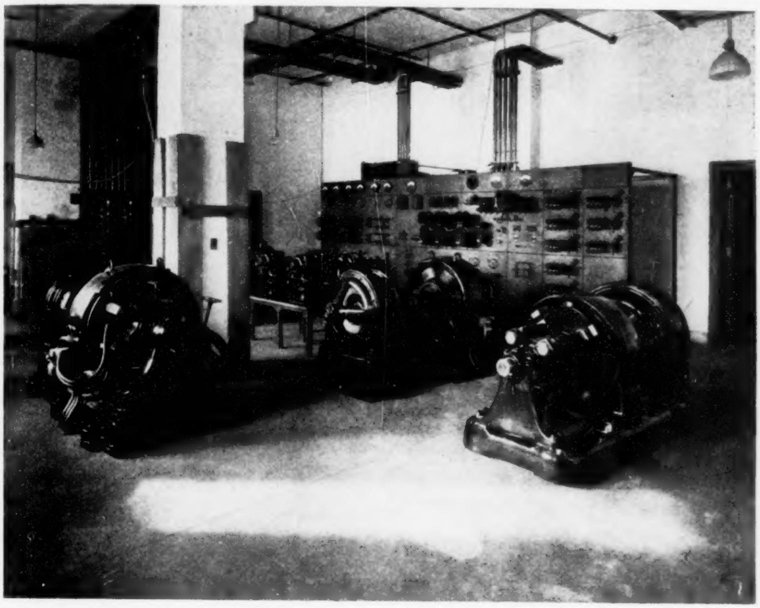


Fig. 15—Power Machine and Control Equipment for Two 10,000-Line Units

Storage batteries (Fig. 16) floating across the current supply bus-bars insure regulation. In addition to stabilizing the voltage and reducing noise interference from the machines and between telephone circuits, the batteries perform the important function of keeping the exchange in operation during interruptions to the commercial power service. Small motor generators furnish current for ringing subscribers' bells and drive commutators supplying various tones and signals. Batteries or machines supply current for operating coin

boxes and pulse machines provide impulses for the operation of certain of the machine switching apparatus.

Whenever practicable, two or more commercial power services from independent generating stations are secured, either of which will keep the office supplied. Where independent generating systems are not available a reserve gas engine supply (Fig. 17) is installed to take the place of the incoming power service, such engine also being equipped for emergency operation on gasoline.

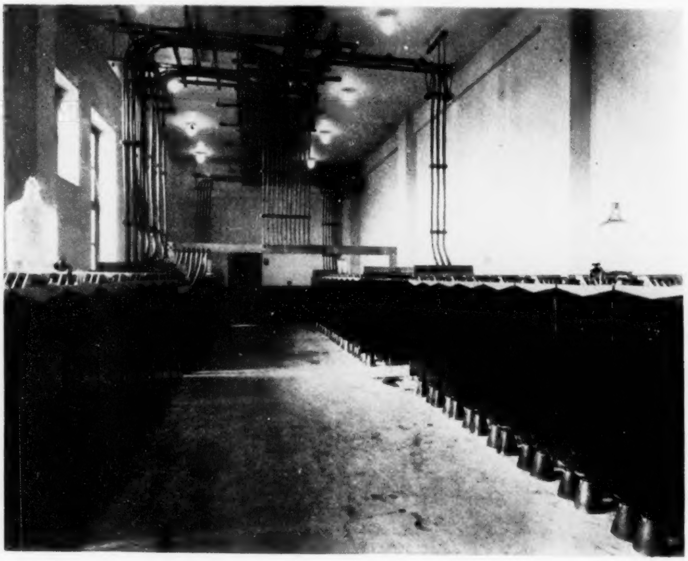


Fig. 16—Battery Room for Two 10,000-Line Units

All of the essential power machines and batteries are provided in duplicate and are arranged to come into action automatically wherever this is necessary to insure continuity of service in the event of loss of power or trouble with any of the power equipment. Alarms are provided to detect variations in battery voltage, blowing of fuses, stopping of machines or any failure of service on all power busses which feed energy to the telephone or signaling circuits. The power plant is thus designed to give an uninterrupted energy supply at all times even when the usual sources of power may have been temporarily discontinued.

Detailed Plan of Operation

The following will give in some detail the plan of operation for handling typical calls between various types of offices in a large metropolitan area such as New York City.

Calls Originating in Machine Switching Offices. Fig. 18 shows schematically the path of a call originating in a machine switching office. The pair of wires of a subscriber's line is attached to one of the sets of fixed terminals in a panel bank appearing before a group of selectors of the type which has been described. By putting fewer

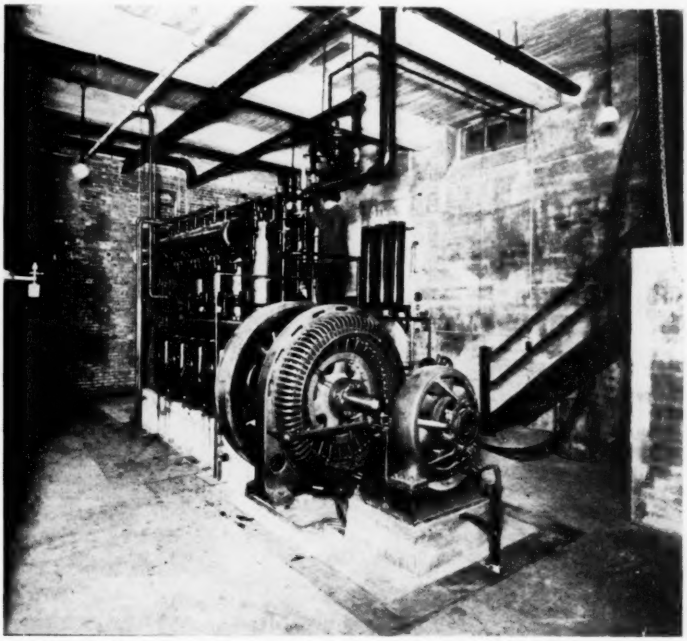


Fig. 17—165 H. P. Gas Engine Generating Set for Emergency Use

lines in these panels and increasing the number of selector brushes, we attain the speed necessary at this stage of the connection. These selectors are called "line finders," since their function is to find calling lines. The terminals correspond to the answering jacks and the selectors to the "A" operators' answering cords of the manual system.

When the subscriber removes his receiver, he closes the circuit of his line, causing a relay at the central office in series with his line, to operate. This relay causes an idle line finder, having access to his line, to trip the proper brush and then move upward to his line. At the same time a sender selector attached to that line finder is choosing,

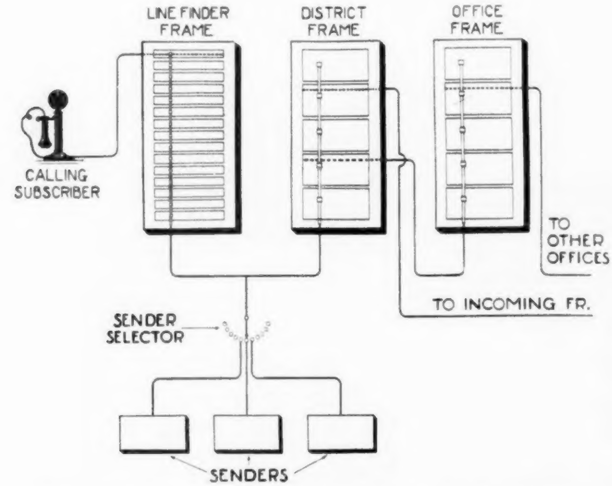


Fig. 18—Diagram of Line Finder, District and Office Frames

out of a common group, an idle sender. The sender selector is a small selector of a type in which the brushes are driven by a magnet over contacts arranged as shown in Fig. 19.

The sender having been attached in this manner to the calling line, a low humming sound, known as the dial tone, is heard by the subscriber, advising him that the mechanism is ready for him to dial. The entire sequence of events just described takes place in a fraction of a second, so that ordinarily the subscriber finds the dial tone when the receiver reaches his ear. The subscriber now dials the required letters of the office name, and the numerals of the called number.

The pulses from the dial come over the subscriber's line through the line finder and sender selector to the sender which records and translates them to control the setting up of the connection. As soon as the connection has been established, the sender is released and is ready to be used for a new call, being kept in use only a few seconds for each call.

The first step in completing the connection is to choose an idle trunk in the proper direction. To the nearby offices there are groups of direct trunks, whereas the more distant offices are reached through tandem centers described later.

The line finder leads to the movable element of a panel selector known as a "district selector." This district selector has capacity for 450 working outgoing trunks, the other 50 trunks being used for control purposes. In a small city 450 trunks would be sufficient to

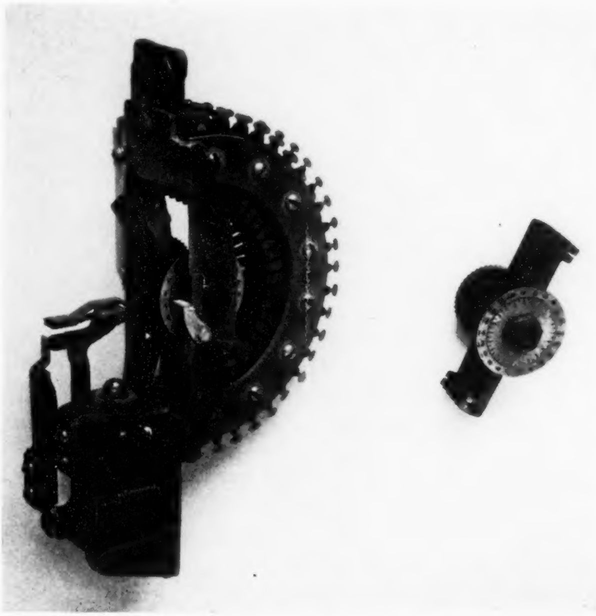


Fig. 19—Sender Selector

reach all points, but in the case of the New York offices 450 outgoing trunks are not sufficient. Accordingly, only a few of the trunk groups outgoing from these offices leave directly from the district selectors. To obtain access to the remaining trunks there are, on every district selector frame, groups of trunks leading to so-called "office selectors." These office selectors are of the panel type and each has a capacity for 450 outgoing trunks.

The path of a call through a district and office selector will now be traced. The district selector starts upward under the control of the

sender. As the district selector moves upward, it produces pulses by means of the brushes which slide over the commutator at the top of the selector. These pulses are transmitted back to the sender, and are there counted. When the sender has counted the number of pulses which indicates to it that the district selector has proceeded to the proper position, the sender opens the fundamental circuit to the selector and causes it to stop. This method of controlling the movment of the selector is termed the reverse control method.

The first selection made chooses the set of brushes to be tripped into engagement with the terminals. Assume, as shown in Fig. 18, the desired trunk appears on the second panel from the bottom. Therefore, the district selector is allowed to make two pulses and is then stopped by the sender. The brush-tripping device is thus set in position to trip the second brush, and the selector is started again by a signal from the sender, which operation completes the process of tripping the brush.

The selector now continues upward, making a pulse for every group of trunks which it passes over, until, having reached the desired group, as indicated by the number of pulses counted by the sender, it is again stopped by the sender at the beginning of this group. The selector is now started again, and this time under its own control, hunts for an idle trunk in the group. Busy trunks are grounded on the third or signaling terminals, whereas idle trunks are open. A testing relay, associated with the selector, keeps the selector moving upward until a trunk with an open third wire is found, whereupon the selector stops, makes connection with this trunk, and renders it busy to other selectors by grounding the signaling strip.

This trunk, as indicated in Fig. 18, leads to an office selector. The same process is repeated by the office selector, under control of the sender, to trip first the proper brush, then choose the proper group, and finally to choose an idle trunk in the group. The connection is now extended to an outgoing trunk. The sender still remains attached to the connection, since it must still control the further setting up of the connection.

The sizes of the working trunk groups on district and office selectors can vary from 5 to 90, depending upon the traffic to be handled.

Calls Between Machine Switching Offices. If the call is for a subscriber in a machine switching office it is completed as shown in Fig. 20. This figure shows a diagram of the apparatus used to connect an incoming full mechanical trunk to a subscriber's line, whether this line is in the originating machine switching office or in another which must be reached over interoffice trunks.

The incoming trunk to the machine switching office terminates on an "incoming selector," which is of the type already described. The machine switching office has a capacity for 10,000 numbers, but the incoming selector has capacity of only 500 trunks, so that the same arrangement is employed as on the district selectors; that is, the incoming selector chooses one of a number of other selectors, called "final selectors," which have access to the subscribers' lines. Since each group of final selectors has access to 500 subscribers, 20

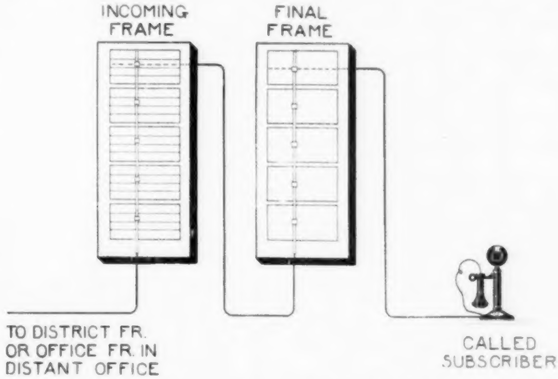


Fig. 20—Diagram of Incoming and Final Frames

groups of finals will be necessary to care for the full 10,000 numbers. On the incoming selector frames, therefore, appear 20 groups of trunks, each group leading to a different frame of final selectors.

The method of selection is the same as described for the district and office selectors; that is, first the incoming selector, under control of the sender in the originating office, trips the proper brush, chooses the proper group, and finally chooses an idle trunk leading to a final selector. The final selector then goes through the process of brush, group, and subscriber's terminal selection. The terminal selection is under the control of the sender which counts line by line in the group of ten, until the desired one is reached. If the called line is idle, it is rung, and the calling subscriber is advised of that fact by hearing the audible ringing signal. If the called line is busy it is not connected, but an intermittent buzz, recognized as the busy signal, is sent back to the calling subscriber. If the called number is that of a P. B. X. having several trunks, the final selector automatically hunts for an idle one. If the final selector, after testing all the P. B. X. trunks finds them all busy, it sends back the busy signal.

As soon as the called line is reached, the sender is dropped from the circuit to be available for another connection. It is not held during the period of ringing, during the time that the busy signal is being given, if the line is busy, or during any part of the period of conversation.

It will be noted that the method of selection is not on a decimal basis. The first selection is to choose one of five brushes on the incoming selector as already explained; that is, we choose that partucular fifth of the terminals in which the called line happens to be, and since 1-5 of 10,000 is 2000, we choose the 2000 group desired. The next selection is by groups of 500, which is again non-decimal. This "translation," as it is called, of the number from the decimal notation, as dialed by the subscriber, into the notation as needed by the selectors, is taken care of very simply in the senders.

Calls from Machine Switching to Manual Offices. Calls from machine switching to manual offices are handled at the manual office on call indicator "B" positions. Fig. 21 shows a diagram of the equipment used to connect such a call to a subscriber in the manual office.

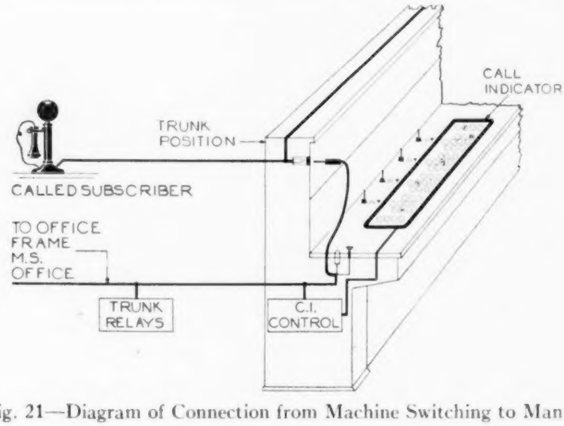


Fig. 21—Diagram of Connection from Machine Switching to Manual

The call progresses through the district and office selector in the same manner as described for the machine switching call, but the trunk which it takes up leads to a call indicator "B" position in the manual office selected. The operator is notified that a call has reached her position by the lighting of a lamp associated with the cord and plug in which the incoming trunk terminates. Upon perceiving this signal, she presses a display key associated with that trunk, and thereupon the called subscriber's number is displayed on a bank

of numbered lamps located on this operator's keyboard. The operator picks up the plug, tests the called line and, if it is found idle, plugs in; or, if it is found busy, she plugs into a special jack which is arranged to send the intermittent busy tone back to the calling subscriber.

The called subscriber's number is displayed in the following manner. Associated with the operator's position, and with her call indicator, is a group of relays. When the display key is depressed, this group of relays is attached to the trunk. The sender which has meanwhile been waiting on the connection, is thereby given a signal, and sends the number called by means of code pulses which are received by the group of relays. These relays, in turn, light the set of lamps on the call indicator corresponding to the digits of the called number, as shown in Figs. 22 and 23. The code pulses employed for sending

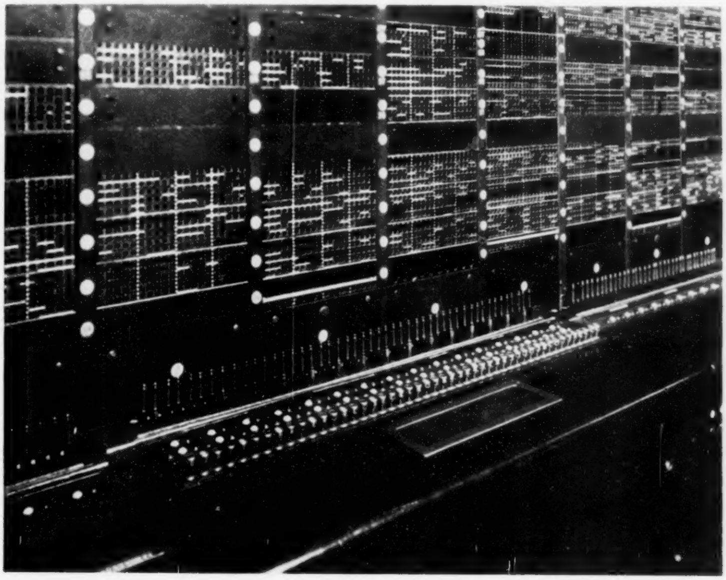


Fig. 22—Incoming Trunk Position in a Manual Office Arranged for Call Indicator Operation

this called number are positive and negative, strong and weak, and are translated by the sender from the decimal dial pulses to this type of pulse to reduce the time required and to simplify the receiving apparatus.

Incoming Calls from Manual to Machine Switching Offices. Calls from manual offices are handled at the machine switching office on the cordless "B" positions. Fig. 24 shows a diagram of the equip-

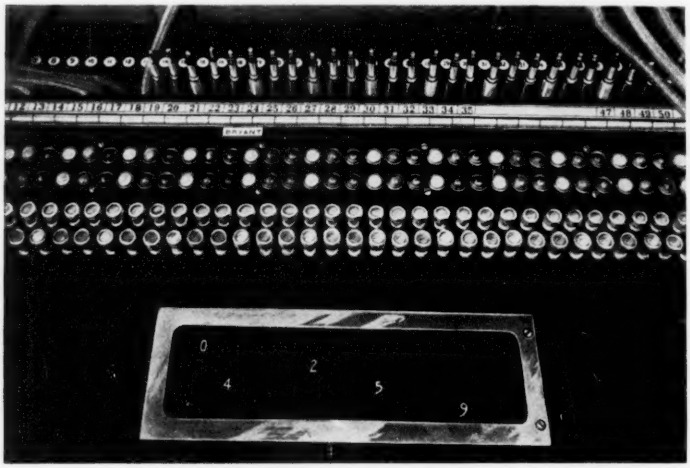


Fig. 23—Call Indicator at an Incoming Trunk Position in a Manual Office

ment used to connect a call originating in a manual office destined for a subscriber in a machine switching office. Such a call is answered by the "A" operator in the manual office in the usual manner. She takes up the call circuit by depressing her call circuit key to the

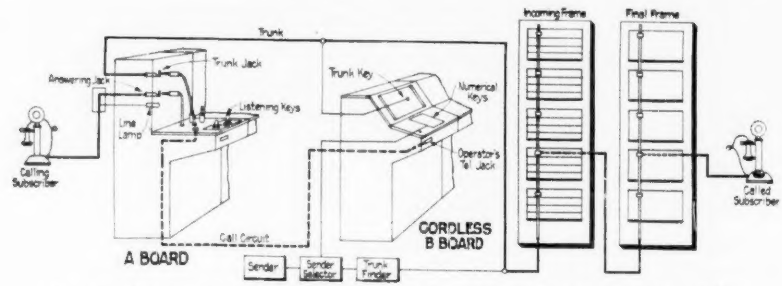


Fig. 24—Diagram of a Connection from a Manual to a Machine Switching Office

machine switching office desired, passes the called subscriber's number, and receives a trunk assignment in exactly the same manner as if the call were going to another manual office. The cordless "B"

operator, upon assigning a trunk, presses the assignment key of that trunk, which temporarily attaches her keyboard to a sender and simultaneously to the incoming trunk which she has assigned. As shown in Fig. 24, the incoming trunk terminates on an incoming selector which has access to final selectors on which the called number appears, in the same manner as described before.

The operator now sets up on her numbered keys the number desired, and this information is transmitted immediately to the sender. These keys, which lock mechanically, are released after a fraction of a second by a magnet controlled by the sender and are ready for the

Fig. 25—Cordless "B" Positions in Machine Switching Office

next call. The "B" operator's sender now controls the incoming and final selectors in the same manner as the subscribers' senders, causing the incoming selector to choose an idle trunk to a final selector having access to the desired group of 500 numbers. The final selector reaches its destination in the manner previously described and, as soon as the line is found, the sender is released.

Fig. 25 shows a line of cordless positions. The section at the left is the cable turning section, having nothing to do with the operation of the board.

Manual Positions Required in Machine Switching Offices. While regular calls between two subscribers will be completed in this system without the aid of operators, certain classes of calls, such as toll calls to suburban points and calls for discontinued or changed numbers, etc., will require the assistance of an operator. Special manual positions are therefore provided in the machine switching office for this service. These positions also care for cases where the subscriber may need the assistance of an operator for other reasons than the above, and are in addition to the cordless "B" positions previously described.

The operators are called "Special Service Operators." The subscriber signals them by dialing "Zero," which on the dial is also marked with the word "Operator." The connection then progresses in the same general manner, through the district and office selectors, as for any originating call. An idle trunk appearing on the office selector leading to an answering jack before the special service operator is chosen and the sender released. Should a subscriber in any local service area dial a subscriber in another area, the sender will automatically route the call to a special service operator.

The special service operator in large areas has before her a number of cord circuits having one end terminating in a cord and plug. She also has upon a keyboard a set of keys similar to those described for the cordless "B" position, except that there are additional strips of keys upon which she can write up an office code. The operator answers the subscriber by inserting one of the plugs in the answering jack and, having ascertained the desires of the subscriber, directs the connection to the proper destination by setting up on her keys the proper numerical code. Senders are furnished for these positions so that, as soon as the information from the keyboard has been registered on the sender, the keys are released and are ready for another call.

The other end of the special service operators' cord circuit terminates in a district selector which, either directly or through other selectors, has access not only to trunks which the subscriber himself might call, but also to trunks leading to more distant offices which he cannot dial directly because they are toll points.

Tandem Operation. There are about 158 central offices in the area shown on the map, Fig. 3. While it is an essential requirement that any subscriber connected to any of these offices be able to reach any subscriber connected to any other office, it is obvious that to furnish trunks from each office direct to every other office would require a great number of long trunks in small groups carrying a very light load most of the time.

In order to eliminate the inefficiency that such an arrangement would entail, it has been the practise in manual operation to handle the traffic from one part of the area to another part of the area over main trunk routes. The collecting and distributing points on these trunk routes are known as "tandem centers," and the plan of operation is known as "tandem operation."

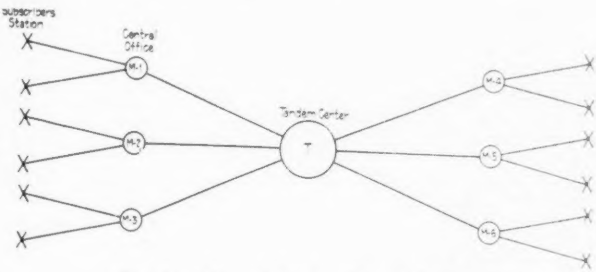


Fig. 26—Typical Tandem Trunking Plan

Fig. 26 shows an arrangement of offices in a typical tandem trunking plan. Offices marked *M* are local offices, either manual or machine switching. The office marked *T* is a tandem office. If a call is originated by a subscriber in office M-1 for a subscriber in offices M-4, M-5, or M-6, to which no direct trunks are provided, the call is routed at office M-1 to trunks terminating at tandem office T. At this point they are connected to trunks leading to the proper office, where the connection is completed to the desired subscriber in the usual manner. Likewise, calls from offices M-2 and M-3 are completed over the same groups of trunks from the tandem office T to offices M-4, M-5, or M-6.

The plan described above is typical of that followed in the New York Metropolitan area for many years, the completion of the call being controlled at the tandem office by operators.

The machine switching system is not only adapted to fit into the existing tandem plan, either when used in the local central office or at the tandem office, but also makes available possibilities for considerably extending the field of usefulness of the tandem system, due to certain advantages in handling calls at tandem points by the use of machinery.

The use of a sender at the machine switching office which is capable of routing a call in any way desired permits locating the selectors which have access to the interoffice trunks at any convenient point either at the originating office or at some distant point. In other

words, the tandem office T shown on Fig. 26 may consist of a group of office selectors such as have been described previously. In this case the trunks from offices M-1, M-2 and M-3 would lead from district selectors in these offices to the office selectors at office T which would select, under control of the sender in the originating office, an idle trunk to office M-4, M-5, or M-6, as desired. At the terminating office the call would be completed through incoming and finals if it is a machine switching office, or call indicator "B" positions if it is a manual office, exactly as described previously.

If the number of points to be reached through the tandem office is greater than the capacity of a group of office selectors, a group of district selectors may be provided at the tandem office which have access to groups of office selectors located at the same office or at some distant point, as described above.

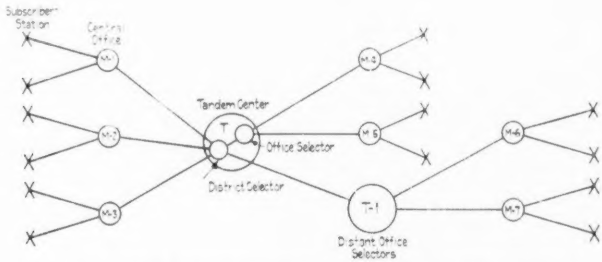


Fig. 27—Tandem Trunking Plan Showing Distant Office Selector

Fig. 27 shows schematically a tandem plan using the above method. Tandem office T is provided with district selectors on which terminate trunks from local office M-1, M-2 and M-3. These selectors have access to office selectors in the same office through which offices M-4 and M-5 are reached, and to office selectors located in the distant tandem office T-1 through which office M-6 and M-7 are reached.

To handle calls at a machine switching tandem office originating from manual offices, operators are required at the tandem office. These operators handle calls in much the same manner as cordless "B" operators in a machine switching office, as already described. The operator receives the desired office name and number from the originating operator over a call circuit and sets it up on her keyboard, which is similar to the cordless "B" board, except that it has office keys in addition to numerical keys. The number is received by a sender which then controls the operation of the selecting mechanism in the

tandem office and other offices through which the call may pass, to the desired local office and subscriber's line.

Many different combinations of the above are possible and are employed when desired.

MAINTENANCE

As will have become apparent from the description already given there is, in the machine switching telephone central office, a large amount of apparatus which, in order to insure service of good quality, must be maintained in proper working condition. Consequently, the subject of maintenance has been very carefully kept in mind throughout the design of the system. For instance, all new pieces of apparatus used in this system have been subjected to the most rigid tests to insure that they will have a satisfactory life and that their margins of adjustment will be adequate.

When maintaining machine switching equipment, the main reliance is placed on preventive measures, so that incipient faults will be detected and corrected before they have got to the point of interfering with service. Ingenious automatic testing arrangements have been designed to aid in this preventive maintenance work. They subject the various circuits in the exchange to routine tests, and are arranged so that they will automatically test all of the circuits, one by one, under conditions more severe than they will ever be called upon to meet in service. In case some feature of any circuit has deteriorated from its normal standard of adjustment—which includes a wide margin—so that it will not meet this severe testing condition, the testing apparatus automatically stops and by supervisory lamps indicates the location of the trouble. An audible alarm is also sounded which notifies the maintenance man responsible that something requiring his attention has been found. The circuit in trouble may still be capable of giving service, but is below the standard set and may soon give service trouble if not corrected.

As applied to the sender, for example, the automatic routine test equipment picks up each sender in turn and puts it through its regular process of operation, under conditions more severe than are encountered in practise. If the sender under test meets the operating conditions without failure, the sender is dropped and the test equipment moves to the next sender. If any trouble develops an alarm is given, which summons the maintenance man who is able to determine by the condition of the apparatus the location of potential trouble.

The operation of the testing equipment may be varied by suitable keys, so that all the features of each sender may be tested once, or so that any one feature of the sender may be tested as many times as desired.

All the equipment in the office occurs in groups, and arrangements are made for readily taking out of service for readjustment any piece of apparatus which may have been found to have potential trouble—the other members of the group continuing to handle the calls.

Application

In the preceding pages there has been briefly described a switching system which meets the exacting and complex requirements of telephone service in the largest cities and in which, so far as is practicable, the various switching operations are performed automatically. Only such operators are required in connection with this system as are necessary for handling special classes of service and certain operations in connection with the interchange of calls between manual and machine switching central offices during the transition period.

Variations in the arrangements which have been described have been developed and are available for use whenever the conditions warrant. An illustration of this is the so-called key indicator, which permits the handling of calls from manual to machine switching offices without the aid of the cordless "B" operators. This is effected by providing the operators in the manual offices with special keys and equipment for controlling directly the selection of the subscriber's line in the machine switching office.

This machine switching system marks a very important advance in a development which began shortly after the telephone was invented, and which has been most vigorously prosecuted by the engineers of the Bell System from then to the present time. Throughout this entire development period the tendency has been to introduce automatic methods and apparatus whenever they gave a better result to the public, or whenever they were attended by an economy of any kind.

How this system works has been briefly explained. What arrangements are provided for handling regular machine switching calls, calls to and from existing manual offices, private branch exchanges, etc., has been described. How the introduction of this system into a telephone network is affected will now be discussed briefly.

Obviously, the problem of introducing machine switching equipment into such an extensive and complex structure as is the telephone

plant of a big city, is a large one. It is impracticable to introduce it all at once. Its introduction must be effected gradually and this is accomplished by using it for growth and such replacements as are necessary, later extending its use as conditions warrant.

The fundamental engineering studies which have to be made and which must precede the manufacture and installation of the equipment for a machine switching office are, in all important respects, the same as those which must precede the manufacture and installation of the equipment in a new manual office. They involve a careful study of the telephone needs of the area, with a view to determining ultimately the quantities of the different kinds of arrangements necessary to give the service. This requires a study of the commercial requirements at the time when the equipment should be cut over and for several years thereafter. Data must be collected as to the probable rates of calling, the average duration of the calls and the amount of trunking to and from other offices.

With these data available, the size and arrangement of the trunk groups on the selector frames, the number, grouping and type of selectors and senders required, and the size of the power plant can be determined. From this the cabling arrangement can be worked out, and suitable floor plans prepared.

Manufacturing specifications can then be prepared in accordance with which the equipment of the office is manufactured and installed. Before the equipment is cut into service, the various arrangements are thoroughly tested individually, and when in proper condition the whole is checked up by making complete operation tests.

If time and space permitted, it would be of interest to discuss the methods of actually cutting the equipment into service, and the comprehensive program which is worked out for the training of the employees who are to handle the equipment and advising the public which is to use it. All these matters are of the utmost importance, and must be carried out systematically in order that there may be no reactions on the general service at the time of the cut-over.

Relations of Carrier and Side-Bands in Radio Transmission[1]

By R. V. L. HARTLEY

SYNOPSIS: This paper discusses generally the characteristics of carrier transmission as applied in radio and in carrier current communication over wires and analyses the factors which affect the faithfulness with which such systems reproduce the signals imparted to them. Modulation is shown to generate two side bands which, with respect to frequency, lie just above and just below the carrier frequency, the frequency width of each side band being the same as the frequency width of the original signals. Upon detection, currents of frequencies corresponding to the difference frequencies between all the possible pairs of component frequencies of the side bands and carrier are produced and, in general, are all found in the received message. It is therefore impossible to transmit messages, either telephone or telegraph, by carrier which will be absolutely free of distortion, but since the amplitude of any particular difference frequency is proportional to the product of the amplitudes of its two generating frequencies the distortion can be reduced below a troublesome value by maintaining the amplitude of the original carrier sufficiently large with respect to the amplitudes of the signal components. The distortion which arises from phase shifts between the component frequencies of the transmitted message and carrier is also considered.

The paper discusses single side-band transmission and carrier suppression with homodyne detection and their various merits are pointed out. Single side band transmission reduces the width of frequency band required for each message. Carrier suppression results in a saving of power, or a more economical expenditure of power, it having been determined that for proper freedom from distortion the power of the carrier component alone, when transmitted, should be rather larger than the peak power in a carrier suppression system. The use of local carrier in homodyne radio telephony assists in frequency selection in the same way as does the heterodyne wave in radio telegraph reception. The same applies also to static interference and, as the object of high power stations is to make the signals large compared with static, there is a gain in concentrating the power in side bands rather than in carrier.

Consideration of distortion arising from phase shifts shows that in homodyne telegraphy distortion can most readily be avoided by transmitting both side-bands, while in telephony these factors favor the transmission of only one side-band. The power of the reproduced signals is twice as great with two side-bands as with one, but there is no choice between one and two side-bands on the basis of the ratio of signals to interference.

The result of using a local detecting frequency which is not exactly equal to the original carrier frequency is discussed, and a balanced detector is described by means of which the distorting effect of the received carrier may be very much reduced. Considering a local carrier which is out of synchronism with the original carrier, it is again found that single side-band transmission is most favorable in telephony, and the transmission of both side-bands is best in telegraphy.—*Editor.*

AS indicated by the title, this paper will discuss some of the phenomena associated with radio transmission in terms of the carrier currents and side-bands into which a modulated wave may

[1] Presented before The Institute of Radio Engineers, New York, December 13, 1922. Printed in the Proceedings for February, 1923, and reprinted here by permission of the Institute.

be resolved. The use of these terms implies a point of view which perhaps is employed less commonly in radio engineering than in some of the other branches of the communication art. For this reason, I shall, at the risk of repeating much that is already in the literature,[2] review such of the fundamentals of this viewpoint as are necessary to an understanding of what is to follow.

Analysis of a Signal Wave

Briefly stated, the point of view is that any signaling wave may be resolved into sustained sinusoidal components, which may be thought of as traversing the system as individual currents and recombining at the receiving end to form the reproduced signal. The possibility of such a resolution has been demonstrated mathematically and the formulas for evaluating the amplitudes and phases of the components are well known. A periodic wave may be expressed as a Fourier series, that is, as the sum of an infinite series of components the frequencies of which may be thought of as harmonics of a fundamental frequency which is equal to the frequency of repetition of the wave. Such a resolution, however, is not directly applicable to the waves employed in communication, for by their very nature they are not periodic. A communication system must be capable of transmitting any individual symbol regardless of what precedes or follows it. We may, however, resolve such an aperiodic wave by the mathematical device of assuming it to be one cycle of a periodic wave in which the interval between successive occurrences of the disturbance in question approaches infinity. The frequency of repetition is then infinitesimal. The fundamental frequency of the Fourier series and the frequency interval between adjacent components are also infinitesimal; that is, the series of discrete lines of the Fourier series spectrum merge into a continuous spectrum. Mathematically this continuous spectrum is represented by the expression

$$F(t)=\int_o^\infty S\ cos\ (qt+\theta)\ dq, \qquad (1)$$

which is known as the Fourier integral. Physically we are to picture this infiinte series of sustained sinusoids as having such amplitudes and phases that the algebraic sum of their instantaneous values is

[2] "Carrier Current Telephony and Telegraphy," by E. H. Colpitts and O. B. Blackwell; "Transactions of the American Institute of Electrical Engineers," volume XL, page 205, 1921. "Application to Radio of Wire Transmission Engineering," by L. Espenschied; presented before The Institute of Radio Engineers, January 23, 1922.

zero for all instants before and after the disturbance in question, and equal to the instantaneous value of the wave thruout its duration. In Fig. 1, curve A represents a telegraph dot, curve B gives the rela-

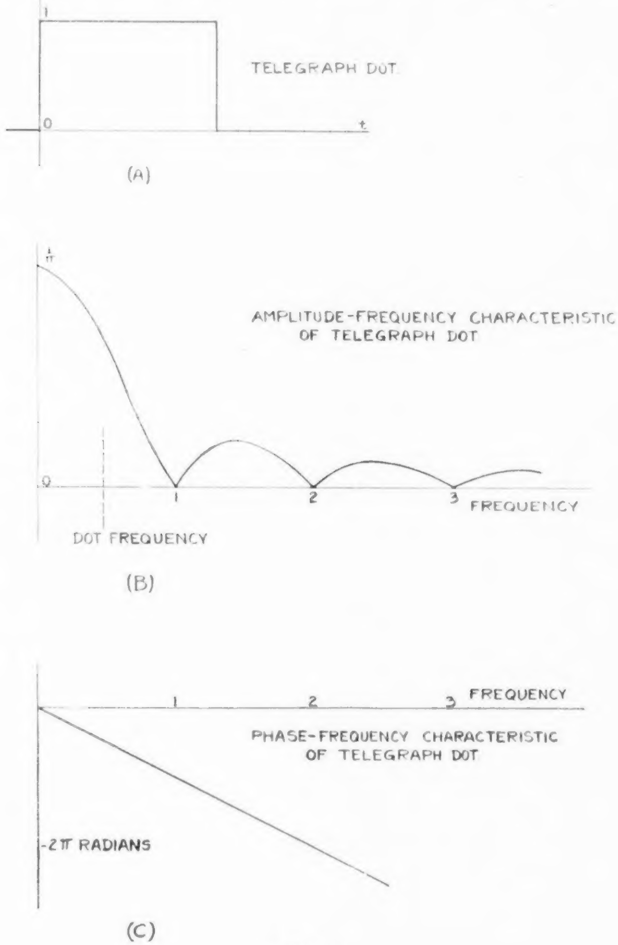


Fig. 1

tive amplitudes, S, of its components plotted against their frequencies, and curve C, their phase, Θ, also as a function of frequency. The so-called "dot frequency" corresponding to a sustained succession of such dots is indicated on curve B.

It is obvious that if either the amplitudes or the phases of the components be distorted, their instantaneous sum will be changed; that is, the wave resulting from their re-combination will be a distorted reproduction of the original wave. Also, those parts of the frequency range in which the amplitude is negligibly small can contribute little to the reproduced wave, and the elimination of all components in those ranges will have little effect on the quality of reproduction. Just what ranges it is essential to retain depends upon the nature of the signal and the standard of reproduction that is set up. What is important for present purposes is the fact that the faithfulness with which a system will reproduce any arbitrary signal disturbance is deducible, in theory at least, from a knowledge of its transmission of sustained single frequencies. By this is meant a knowledge of how the relation, both in amplitude and phase, between the input and output sinusoidal wave varies as the frequency of the wave is progressively varied thruout the frequency range.

Analysis of a Modulated Wave

Let us assume now that a radio system is called on to transmit such a signal wave, $F(t)$, which may be either a telephone or a telegraph signal. If, as is commonly assumed, the modulator causes the amplitude of the carrier wave, $C \cos p\,t$, to be varied in accordance with the signal, the resulting modulated wave may be expressed as

$$m = C\,[1 + k\,F(t)]\cos pt, \tag{2}$$

where k is a factor which measures the so-called degree of modulation. If the largest negative value of $k\,F(t)$ is just equal to unity, so that the instantaneous amplitude of the carrier wave just falls to zero, the modulation is said to be complete. The significance of complete modulation will be discussed later.

Now let us resolve the signal wave into its infinite series of components, each of the form $S \cos(qt+\theta)$, where S and θ vary with the frequency $\frac{q}{2\pi}$. Neglecting non-essential frequencies, q may be considered to cover a range from q_1 to q_2. If this value of $F(t)$ be substituted in (2) we get

$$m = C\cos pt + k\,C\cos pt\int_{q1}^{q2} S\cos(qt+\theta)\,dq. \tag{3}$$

The first term, which is independent of the signal, represents a component having the carrier frequency, $\frac{p}{2\pi}$. The second term represents an infinite series of terms each derived from only one component of

the signal. Hence each component of the signal is represented in the modulated wave by an expression of the form,

$$k\,C\,S\cos(qt+\theta)\;\cos pt=\tfrac{1}{2}k\,C\,S\;\{\cos[(p+q)t+\theta]+\cos[(p-q)t-\theta]\}.\quad(4)$$

This represents two sinusoidal components, the frequencies of which differ from that of the carrier by the frequency of the particular signal component. The similar expressions for the other signal components each yield a pair of components similarly placed with reference to the carrier. All of these taken together form a pair of spectra or frequency bands extending on either side from the carrier frequency in the same way that the spectrum of the signal extends from zero frequency. These bands of frequencies are spoken of as "side-bands" and the component currents of these frequencies as "side-band currents," or, more often, simply as "side-bands." The side-band which extends upward in frequency from the carrier is called the "upper side-band," and the other, which extends downward, the "lower side-band."

The form of these side-bands is shown schematically in Fig. 2, where purely arbitrary curves are used to represent the amplitudes and

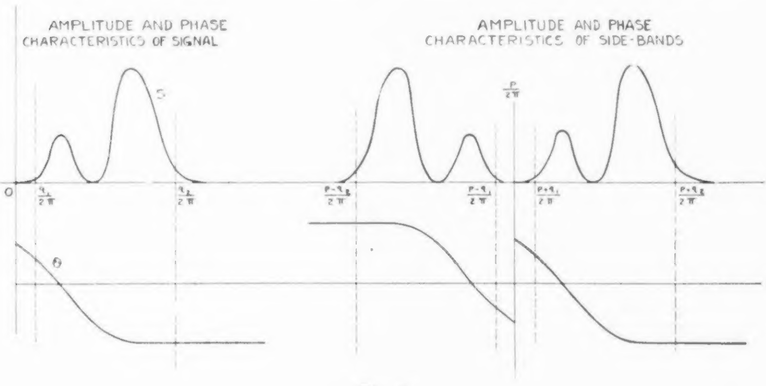


Fig. 2

phases of the signal components over a limited frequency range. It will be seen that the corresponding curves for the upper side-band are derived from these by displacing them along the frequency axis by the amount of the carrier frequency. The amplitude curve of the lower side-band is derived by inverting that of the upper with respect to the carrier frequency. For the phase curve of the lower side-band that of the upper is to be similarly inverted and also reversed in sign.

The actual magnitude of the side-band currents relative to the carrier depends on the degree of modulation, k, of equation (2). For commercial telephony the limits of the essential band may be taken roughly as 200 and 2,000 cycles. If high quality speech or music is to be transmitted, a wider band is required. For telegraphy the band width required varies widely with the speed of sending and the type of apparatus used. In general, it is desirable to preserve very low frequencies, which means that the two side-bands practically meet at the carrier frequency.

Reproduction of the Signal Wave

Having arrived at a picture of the modulated wave as given by equation (4), we shall first discuss the reproduction of the signal from this as it stands, and then consider the effect on this reproduction of various modifications to which the modulated wave may be subjected before or during the process of detection. While any device in which the current-voltage characteristic is non-linear may be used as a detector, the operation of the vacuum tube lends itself to analysis because of its approximation to a parabolic current-voltage relation. That is, we may write,

$$i = a_0 + a_1 v + a_2 v^2, \qquad (5)$$

where v is the voltage impressed on the grid, in this case the modulated wave, and i is the resulting current. As the first term is independent of v and the second represents simple amplification, detection [3] can result only from the third term, $a_2 v^2$. Since a_2 multiplies all components of v^2 alike, we may neglect it and simply consider the square of the expression for the modulated wave. This results in a series of terms which are the squares of the individual components and another which are their products taken in pairs. Since

$$\cos^2 x = \tfrac{1}{2}\,(1 + \cos 2x), \qquad (6)$$

the square terms will yield only direct current, and currents of approximately twice the carrier frequency. The product terms, each of which contains the product of two cosines, may, as in the case of the modulated wave above, be transformed into the sum of two cosine terms the frequencies of which are respectively the sum and difference of the component frequencies. Of these only the difference frequencies can lie in the range of the original signal. In other words, we may think of the reproduced wave as made up of the sum of all the

[3] In practice this parabolic law seldom holds strictly, and secondary contributions are made to the detected wave by terms of higher power.

heterodyne beat notes resulting from all the pairs of component sinusoids of the modulated wave.

The carrier component, $C \cos pt$, beating with a component of the upper side-band, $\frac{1}{2} k C S \cos [(p+q) t + \theta]$, equation (4), gives the beat note or reproduced component,

$$r_+ = \frac{1}{2} k C^2 S \cos (qt+\theta), \tag{7}$$

which is identical in frequency and phase with the corresponding component of the signal, and has an amplitude proportional to that of the signal component. Exactly the same expression results from beating the carrier and the corresponding component of the lower side-band. These two low frequency components, being in phase, add directly to give

$$r = k C^2 S \cos (qt+\theta) \tag{8}$$

as the reproduced component. As the factor $k C^2$ is independent of q, all of the signal components are reproduced with the same relative amplitudes and phases, as in the original signal. Their sum is therefore $k C^2 F(t)$, and the signal is accurately reproduced.

However, there are still other components of the modulated wave to be considered. Every pair of components in one side-band beat to give the difference of their frequencies, which is also the difference of the corresponding signal components. The corresponding pair of components of the other side-band yield an identical component and the two add in phase. Similarly every component of one side-band beats with every component of the other, giving in each case the sum of two component frequencies of the signal wave. Like the difference frequencies, each of these sum frequencies is produced twice. The combination of the components of the two side-bands which were derived from the same signal component yields a component of twice the frequency of the signal component. The addition of these extraneous components serves to distort the reproduced wave in a manner quite similar to that of external interference. It is of interest therefore to consider the magnitude of these distorting components relative to the reproduced signal. The product of two side-band components of amplitudes $\frac{1}{2} k C S$ and $\frac{1}{2} k C S'$, equation (4), gives as the amplitude of one of the two components of the difference frequency, $\frac{q-q'}{2\pi}$, $\frac{1}{4} k^2 C^2 S S'$. Comparing this with the amplitude, $\frac{1}{2} k C^2 S$, equation (7), of one of the two reproduced signal components of frequency $\frac{q}{2\pi}$, the ratio of the undesired to desired component is found to be $\frac{1}{2} k S'$. It is evident that this type of distortion

increases with the degree of modulation, k, or, as will be discussed more fully later, with the ratio of carrier to side-band.[4]

Single Side-Band Transmission

So far it has been assumed that the wave applied to the detector is identical with that produced by the modulator, a condition seldom encountered in practice. For, in addition to the undesired modifications which the modulated wave undergoes because the transmission characteristics of practical circuits are not ideal, there are other changes which when properly made yield distinct advantages. These intentional changes will be discussed first.

It will be remembered that any component of the signal can be reproduced by the combination of the carrier with either side-band. Hence it is unnecessary to transmit both side-bands. Suitably designed electrical filters make it possible to transmit one side-band and effectively suppress the other.[5] This makes possible a very great saving in the frequency range required per channel. It is of particular importance for long wave radio telephone transmission where the width of a single side-band is so large a fraction of the total frequency range available that the number of independent channels is at best very limited. The intensive development of a limited frequency range by the use of single side-band transmission has probably progressed farthest in connection with carrier telephony over wires. Here commercial service is being given over circuits on which the carrier currents of adjacent channels are separated by only 3,000 cycles. It is obvious that the transmission of both bands would nearly double this separation, thereby halving the number of channels per circuit. There is, of course, no reason why similar savings may not be effected in the field of radio transmission. In addition to this major advantage there is an incidental improvement in the quality of reproduction, for the distorted components resulting from beats between components of the two side-bands, that is, the sums of the signal frequencies, are eliminated.

Carrier Suppression and Homodyne Reception

The other important modification has to do with the so-called "unmodulated" component of carrier frequency, $C \cos pt$, in equation

[4] A similar form of distortion generally occurs in modulation, resulting in new components being produced in the frequency range of the side-band.

[5] For a description of such filters see the Colpitts and Blackwell paper referred to above.

(3). As already pointed out, good signal reproduction requires that at the detector this shall not be too small relative to the side-bands. However, it is merely a continuous alternating current, and does not itself partake of the signal variations. It is therefore immaterial whether it is transmitted from the modulator or is supplied to the detector by a local source such as an oscillator. The elimination of this component from the modulated wave at the sending station is spoken of as "carrier suppression," and its re-introduction at the receiving end as "homodyne" or "zero beat" reception. The term homodyne implies supplying the same wave as distinguished from heterodyne, meaning another. Zero beat refers to the bringing of the local carrier into synchronism with the sending carrier by reducing the beat note between them to zero frequency. While homodyne reception is essential to carrier suppression, the reverse is not true. The reception of an ordinary modulated wave may sometimes be improved by the addition of carrier at the receiving end.

The primary advantage of carrier suppression lies in the saving of sending power which it makes possible, or, what is equivalent, the increase in range made possible when all the power of a given station is utilized in the side-band. Of the various ways in which this suppression may be accomplished, the simplest is by the use of a so-called balanced modulator as shown schematically in Fig. 3. Carrier fre-

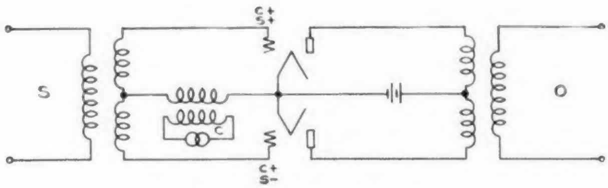


Fig. 3—Balanced modulator

quency from the source C is applied to the grids of two vacuum tubes in the same phase, while signal currents, indicated at S, are applied to the two in opposite phase. The two plate circuits are differentially connected with a common output circuit. In the absence of signaling current the amplified carrier frequency currents from the two tubes neutralize each other and nothing is transmitted. With the application of signaling current one grid is raised in potential and the other lowered, with the result that more radio frequency is developed by the first tube than by the second and the excess appears at O. The magnitude of this radio frequency current is proportional to the instantaneous value of the signaling current. Upon reversal of the

direction of the signaling current the effect of the second tube predominates and radio frequency is again transmitted, this time with the phase reversed, owing to the differential connection. The wave form of a signaling current and the resulting output current are roughly as shown in Fig. 4, *A* and *B*. If this output be amplified for trans-

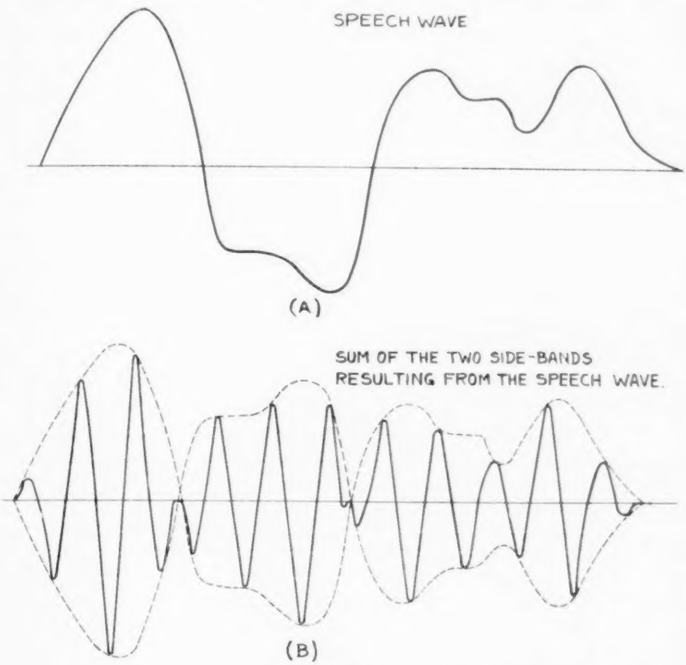


Fig. 4.

mission there will be no load on the amplifier and antenna except during actual speech, when it will be proportional to the intensity of the speech.

That these intermittent pulses of carrier frequency produced by a balanced modulator are equivalent to a modulated wave from which the carrier frequency component has been removed, may be easily shown. Consider a single sinusoidal component $S \cos (q\,t+\Theta)$, of the signaling wave which is applied to the balanced modulator. The resulting output current is a wave of carrier frequency, the amplitude of which is proportional to C and to S and varies cyclically with a frequency $\frac{q}{2\pi}$ between the values $+K\ C\ S$ and $-K\ C\ S$, where K is a

constant of proportionality and the negative amplitude indicates a reversal of phase during half of the audio frequency cycle. Such a variation may be represented by the expression

$$i = K\ C\ S \cos\ (q\ t + \theta)\ \cos\ p\ t. \qquad (9)$$

Taking the sum of these expressions for all the components of the signal gives the second term of equation (3) which was shown to represent the side-bands.

In estimating the power saved by carrier suppression the comparison should be made with a system transmitting the carrier which has been so adjusted that the power is used to the best advantage. So far as a single signal component is concerned this would call for making the carrier and side-band equal, as their product would then be a maximum. This, however, would imply that the distorting currents from the interaction of two side-band components would be as large as the signal currents themselves. That is to say, quality considerations require that the major part of the transmitted power be in the carrier component. Quantitative data on the relation between the ratio of carrier to side-band and the quality of transmission has been secured in the laboratories of the Western Electric Company, and it is hoped it will be published in the near future. Briefly, the results indicate that the good quality which is obtained when the carrier component is large falls off very rapidly as the magnitude of the carrier component is reduced so as to approach that of the side-band, the latter being measured when a sustained "ah" sound is used as the signal. Under these conditions the side-band is sustained at a value about equal to the maximum occurring in ordinary speech. That is to say, even the peak power in a carrier suppression system is less than the carrier component alone in an ordinary system adjusted to give the same side-band. From these considerations it appears that there has been a tendency to attach undue significance to "complete modulation," as a more or less unique and ideal condition of operation. For nothing revolutionary occurs as the carrier is decreased thru the value corresponding to that condition. The distortion due to interaction between the side-bands is present for larger values of carrier and continues to increase progressively for smaller values. The exact degree of modulation to be permitted therefore depends upon the standard of quality to be met. In a carrier suppression system the degree of modulation, k, approaches infinity more or less closely depending on the completeness of the suppression.

In addition to making possible the use of carrier suppression, homodyne reception presents other advantages. It furnishes a ready

means of increasing the intensity of the reproduced signal, since this is proportional to the carrier component at the receiver as well as to the side-band. Also, by making the carrier large, k is made very small and the distorting currents due to interaction of the side-bands become negligible. The use of a large local carrier in homodyne radio telephony assists in frequency selection in the same way as does the hetrodyne wave in radio telegraph reception. Suppose an interfering message is separated from the desired one by only a few thousand cycles and so is not entirely suppressed by the receiving selective circuits. Currents of voice frequency can be reproduced from its side-bands only by interaction with its own carrier, and hence they will be small compared with those of the desired message, which are proportional to the local carrier. On the other hand, the large currents due to the interfering message and local carrier will all have frequencies above the voice range, and so can be suppressed by selective circuits in the output of the detector.

The same general reasoning applies also to static interference. Appreciable interfering currents of signal frequency can result only from those components of the static wave which lie in the frequency range of the side-bands. Moreover, they will bear the same ratio to the signal currents as do the static components to the side-band components. We may conclude, then, that when means are provided for eliminating all of the static except that which is inherently inseparable from the signal, the disturbing effect of the residue is determined solely by the relative magnitude of the *side-band* components and the static components which lie in the same frequency range. As the object of high power stations is to make the signals large compared with the static, the importance of concentrating the power in the side-bands rather than in the carrier is obvious.

Effect of Radio Distortion

Let us pass now from the intentional modifications of a modulated wave and consider the effects of unintentional distortions. Limiting our attention first to systems in which the carrier is transmitted, we have to consider the effect of distortion such as might be introduced by the sending and receiving circuits and the transmitting medium. Assuming the characteristics of these to be known in terms of their transmission of sinusoidal components of various radio frequencies, we wish to determine their effect on the amplitudes and phases of the components of the reproduced wave. We shall assume the current-voltage relations in the transmission system to be linear, so

that no new frequencies are introduced. Then any possible distortion in the modulated wave may be represented by assigning the proper amplitudes and phases to all of the components. Corresponding to a single component of the signal we may write for the received wave

$$m = B \cos (pt-\phi) + B_+ \cos [(p+q)\, t+\theta-\phi_+] + B_- \cos [(p-q)\, t-\theta-\phi_-] \qquad (10)$$

where the amplitude, B, and phase lag, ϕ, may vary in any arbitrary manner for the different components of the modulated wave. We shall assume that B is always large enough compared with B_+ and B_- that the interaction between the side-band components may be neglected. It will be seen that the single frequency components reproduced from the two side-bands are not in general equal nor in phase and may either aid or tend to neutralize each other. They will be of the form,

$$r = B \{ B_+ \cos [q\, t+\theta-(\phi_+ -\phi)] + B_- \cos [q\, t+\theta-(\phi-\phi_-)] \}. \qquad (11)$$

Taking the resultant of these two gives as the component of the reproduced wave,

$$r = R \cos (qt+\theta-\Psi) \qquad (12)$$

where

$$R = B\sqrt{B^2_+ + B^2_- + 2\, B_+ B_- \cos [(\phi_+ -\phi)-(\phi-\phi_-)]} \qquad (13)$$

$$\tan \Psi = \frac{B_+ \sin (\phi_+ -\phi) + B_- \sin (\phi-\phi_-)}{B_+ \cos (\phi_+ -\phi) + B_- \cos (\phi-\phi_-)}. \qquad (14)$$

It is evident that both the amplitude, R, and the phase shift, Ψ, of the reproduced component depend upon both the amplitudes and phases of the corresponding components of both side-bands and on the phase of the carrier. The amplitude depends also on the amplitude B of the carrier, but as variations in this affect all components alike, they do not alter the wave form of the reproduced signal, but only its magnitude.

The expressions for the reproduced wave become much simpler for a system in which one side-band, say the lower, is suppressed. Then

$$B_- = 0 \qquad (15)$$

and equations (13) and (14) reduce to

$$R = B\, B_+ \qquad (16)$$

$$\Psi = \phi_+ - \phi. \qquad (17)$$

The amplitude of the reproduced component is independent of the

phases in the modulated wave, and is proportional to the amplitude of the side-band component. Hence amplitude distortion of the reproduced wave can result only from unequal transmission of the different component frequencies of the side-band. The change in

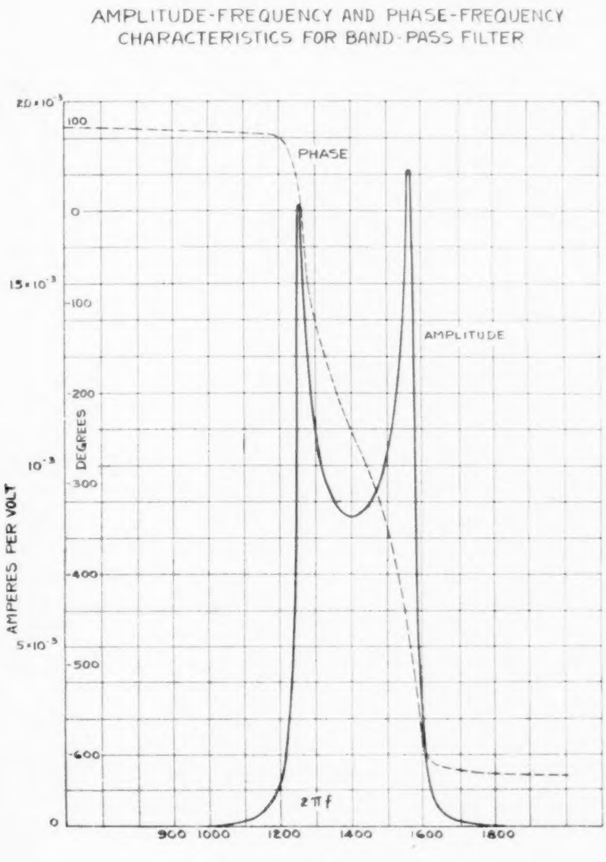


Fig. 5

the amplitude curve for the signal will be identical with that of the side-band. The phase shift, Ψ is independent of amplitude distortion of the modulated wave. It is equal to the difference between the phase lags of the side-band component and the carrier. Fortu-

nately the quality of telephone reproduction is not seriously impaired by shifting the phases of the various components by even as much as several cycles. In telegraphy, however, the shape of the signal current which operates the relay depends very much on the preservation of the proper phase relations of the components, and the entire nature of the signal may be changed by phase shifts of even a fraction of a cycle.

It is worth while then to examine some of the phase shifts which are likely to occur in practice. Transmission of a sinusoidal wave thru the free ether involves a phase lag proportional to the distance and to the frequency. Hence the phase lags, ϕ_+ and ϕ, due to this cause, will be proportional to $p+q$ and p respectively, and their difference, Ψ, will be proportional to q. Replacing Ψ by $h\,q$ in equation (12) and regrouping terms gives

$$r = R\cos\left[q(t-h)+\theta\right]. \tag{18}$$

By displacing the origin of time by h this becomes identical with the original signal component. Also, since h is independent of q, the same time shift brings all the components into agreement; that is, a phase shift proportional to the frequency does not distort the wave, but merely delays it by the corresponding time of transmission.[6] In considering the terminal circuits then it is only the departure of their phase lag versus frequency curve from a straight line that need be considered as a source of distortion. It is of interest to note here that for most filters this relation is approximately linear thruout the range of free transmission. The actual curves for a particular band filter are shown in Fig. 5, where there is plotted against frequency the relation of the amplitude and phase of the current at the third section of an infinite filter to those of the voltage applied to the first section. It will be noticed how the phase curve departs abruptly from a straight line at the edges of the band where the sudden drop in the amplitude curve occurs. Similarly Fig. 6 shows how in the current-voltage relation of a simple resonant circuit, the distortion of phase and of amplitude occur together.

In case both side-bands are transmitted, a simple relation is found if the distortion is symmetrical with respect to the carrier frequency. By this is meant that, however, the different components are distorted relative to each other, for every signal component the two corresponding side-band components are equal in amplitude and are

[6] For a fuller discussion of this point see a paper by T. C. Fry on "Theorie des binauralen Hörens nebst einer Erklärung der empirischen Hornbostel-Wertheimerschen Konstanten," "Physikalische Zeitschrift," 23, page 273, 1922.

shifted in phase relative to the carrier by the same amount; that is to say, for every value of q considered separately,

$$B_{+} = B_{-} = B_{\pm} \tag{19}$$

$$\phi_{+} - \phi = \phi - \phi_{-} = \delta. \tag{20}$$

Then

$$R = 2\, B\, B_{\pm} \tag{21}$$

$$\Psi = \delta. \tag{22}$$

The same considerations as to distortion of the side-band apply here as for the single side-band. There is one point, however, of some practical significance. As is evident from Fig. 6, the charac-

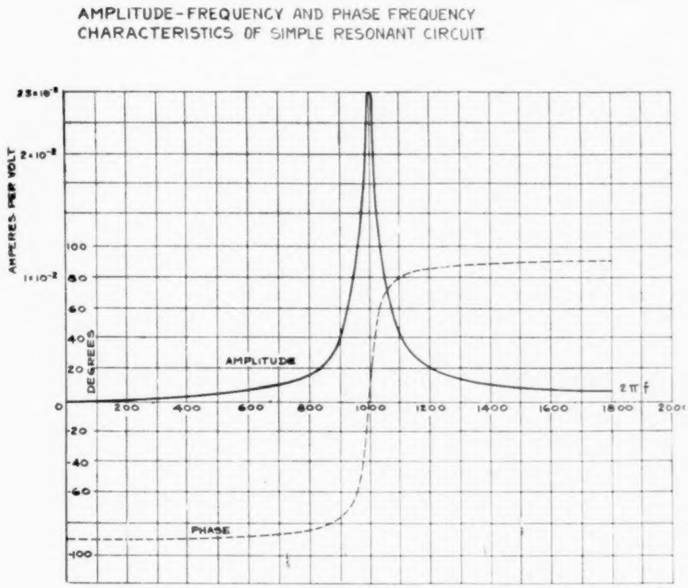


Fig. 6

teristics of a resonant circuit come very close to satisfying these symmetrical conditions if the carrier coincides with its resonance frequency. Its effect on the amplitude and phase curves of the reproduced signal may therefore be derived independently from the amplitude and phase curves of the tuned circuit. If, however, the

circuit be detuned, this symmetry is upset and we are forced to the complicated relations of equations (13) and (14), from which it is difficult to draw general conclusions.

An interesting case of unsymmetrical transmission is that in which one side-band is only partially suppressed owing to insufficient selectivity. Let us assume that the upper side-band is transmitted without distortion. Then for a given amplitude of the lower side-band its effect on the amplitude curve of the reproduced signal will be worst when the phase relations are such that for some frequencies it aids the upper side-band and for others it opposes. The greatest fractional change in the amplitude of any one signal component due to the presence of the lower side-band occurs when the two oppose. It is then reduced in the ratio

$$\frac{B_+ - B_-}{B_+} = 1 - \frac{B_-}{B_+}. \tag{23}$$

Thus, if the lower side-band component were a tenth of the upper, the most it could do would be to change the amplitude of the signal component by a tenth. Such a change would have little effect on telephone quality, particularly as it would have this maximum value at only a few frequencies. The case of telegraphy is rather different. Here the two side-bands lie so close together that it is practically impossible to separate them by radio selective circuits, and even when, as in wire transmission, so low a carrier frequency is used that the two may be separated fairly well by filters, the side-bands corresponding to the lower signal components differ so little in frequency that even a sharp filter does not produce very great discrimination between them. This, coupled with the fact that the phase shift of a filter ceases to be linear near the edge of the transmitted band, leads to very considerable amplitude distortion.

The effect of the unsuppressed side-band on the phases of the reproduced components is also rather complicated. Equation (14) shows that it is a maximum when

$$\begin{aligned} \phi_+ - \phi &= 0 \\ \phi - \phi_- &= 90^\circ, \end{aligned} \tag{24}$$

in which case the presence of the lower side-band changes *tan* Ψ from 0 to $\frac{B_-}{B_+}$. As in the case of amplitude, this effect is unimportant in telephony, but would need to be considered in telegraphy.

Phase of the Local Carrier

Coming now to homodyne reception, the important new factor to be considered is the fact that the carrier component is now perfectly arbitrary in amplitude and phase. This is true even tho the sending carrier is not suppressed, for, by suitably choosing the local carrier, the resultant of the two may be given any desired value. Since the amplitude of the carrier affects only the magnitude of the reproduced signal as a whole, we need consider here only the effect of arbitrary values of its phase, ϕ. For simplicity we shall assume that the modulated wave reaches the receiver unchanged except for the phase lags involved in undistorted transmission. Let us designate by ϕ_1 the phase lag of the carrier which is received from the transmitting station or would be received if it were not suppressed. Then

$$\phi=\phi_1+\eta, \tag{25}$$

where η may be regarded as the phase displacement of the local carrier.

Consider first a system in which one side-band is suppressed. From equation (16), the amplitudes of the reproduced signal componants are independent of the phase of the carrier. From equation (17), the phase,

$$\Psi=\phi_+-\phi_1-\eta. \tag{26}$$

But $\phi_+ - \phi_1$ represents only the phase shifts of undistorted transmission; that is, the delay suffered by the signal as a whole. Hence the net result is that all components have their phases shifted by the same amount; namely, the phase displacement of the carrier, which can never be more than a single cycle. For telephony this is of no practical importance, but it is evident that in a telegraph system using side-band suppression and homodyne reception the phase of the local carrier would have to be very carefully controlled.

Consider now the case of homodyne reception of both side-bands received without distortion; that is,

$$B_+=B_-=B_\pm \tag{27}$$

and

$$\phi_+-\phi_1=\phi_1-\phi_-=h\,q. \tag{28}$$

From these relations and equations (13) and (14) we get

$$R=2\,B\,B_\pm \cos\,\eta. \tag{29}$$

$$\Psi=h\,q. \tag{30}$$

This shows that the amplitude of every component varies as $\cos\,\eta$; that is, when the local carrier is in phase or 180° out of phase with the received carrier the reproduced components are maximum; for inter-

mediate values they decrease, becoming zero when the two carriers are in phase quadrature. The phase lag, $h\,q$, is that due to transmission alone; that is, the phases of the reproduced components are independent of the phase of the local carrier. Since the phase of the local carrier affects only the amplitudes and affects these the same for all components, it does not alter the wave form of the reproduced signal, but does affect its magnitude very materially.

Thus in a carrier telephone system fluctuations in the phase of the local carrier are much more serious when both side-bands are transmitted than when one is suppressed, the only effect then being an unimportant phase distortion. In a carrier telegraph system, however, the amplitude fluctuations which occur when both side-bands are transmitted may not be particularly troublesome since telegraph receiving apparatus is designed to operate over quite a range of signal intensity. The phase distortion occurring in single side-band transmission is however serious. It may perhaps be considered fortunate that the requirements as to phase regulation are least severe in telephony with a single side-band and in telegraphy with both side-bands, since these modes of operation appear on other grounds to be the most practical for the two cases.

In comparing single and double side-band transmission it is interesting to note that for equal sending power, the power of the reproduced signal component is twice as great with two side-bands as with one. However, the power of the same frequency resulting from static is also twice as great, so that the ratio of signal to interference is the same in both cases. To show this, let B_1 be the amplitude of a component of the single side-band and B_2 that of each of the corresponding components of the double side-band. Then equality of power gives

$$B_1^2 = 2\,B_2^2. \tag{31}$$

For the single side-band the amplitude of the reproduced component is

$$R_1 = g\,B_1, \tag{32}$$

where g is a constant of proportionality. The power,

$$P_1 = \tfrac{1}{2}\,g^2\,B_1^2 = g^2\,B_2^2. \tag{33}$$

(The resistance is here omitted as it is assumed constant thruout.) For the double side-band, since the two components are in phase, the resultant amplitude,

$$R_2 = 2\,g\,B_2, \tag{34}$$

and the power,

$$P_2 = 2\,g^2\,B_2^2 = 2\,P_1. \tag{35}$$

If the static be assumed to approximate an impulse, the amplitudes of all its components will be sensibly the same. If we call this amplitude S, then, in the case where the receiving circuit admits only one side-band, the amplitude of the reproduced interfering current of the frequency of the signal component is

$$I_1 = g\, S, \tag{36}$$

and its power,

$$W_1 = \tfrac{1}{2}\, g^2\, S^2. \tag{37}$$

With both side-bands this interfering current is made up of two equal components derived from the static components of frequencies $\frac{p+q}{2\pi}$ and $\frac{p-q}{2\pi}$ respectively. The phase difference ϵ between these two will be accidental, so for any one case the resultant amplitude,

$$I_2 = 2\, g\, S \cos \epsilon, \tag{38}$$

and the power,

$$W_2 = 2\, g^2\, S^2 \cos^2 \epsilon. \tag{39}$$

As all values of ϵ are equally probable, we may average W_2 with respect to ϵ, whence

$$\overline{W}_2 = g^2\, S^2 = 2\, W_1. \tag{40}$$

There is then no choice between one and two side-bands on the basis of the ratio of signal to interference. With a single side-band, the major advantage of economy in frequency range is secured at the expense of the minor disadvantage that to give the same response the amplification of the receiving set must be greater by a factor of two in power, or about three miles of standard cable.

Use of Non-synchronous Local Carrier

In practice, however, unless the receiving carrier frequency is controlled by the same source as the sending carrier, it is rather difficult to maintain even the frequencies alike, to say nothing of the phases. Let us suppose that the local carrier is out of synchronism by a small amount, n. Consider first the simplest case where the carrier is suppressed and one side-band only is transmitted. The local carrier beating with each component of this side-band gives a component of normal amplitude, but of a frequency differing from that of the original signal component by n. That is, all the frequencies of the speech are raised or lowered by the same amount, n. This must alter

the wave form very decidedly, but the surprising thing is that in telephony the intelligibility is not seriously affected when the difference is made as much as fifty cycles or so. The apparent pitch of the voice changes, of course, as n is varied.

If the carrier is transmitted, either intentionally or thru incomplete suppression, the situation is less favorable to asynchronous reception. The two carriers then beat together, giving a component of frequency n which may be troublesome if the received carrier is large. However, its frequency is generally below the voice range, and so it can be suppressed by a filter in the detector output. In addition the received carrier beating with the side-bands gives the components of the original signal. These are superposed on the displaced speech from the local carrier, the corresponding components of the two differing in frequency by n. As a result, the two sounds beat together just as two tuning forks would. For very little differences in frequency a periodic rise and fall in intensity is heard. When the difference is increased so that the individual beats can no longer be distinguished, a sensation of roughness results. And when the difference is made still greater the two waves may be heard as separate sounds of noticeably different pitch. The prominence of this beating effect depends, of course, upon the relative magnitude of the two carriers, since the two sets of speech currents are in the same ratio as the two carriers.

This effect of the received carrier may be very much reduced, and in the ideal case entirely eliminated, by the use of a balanced detector similar in structure to the balanced modulator of Fig. 3. It can be shown that with such a circuit the combination frequencies resulting from any two components applied at S are neutralized in the output circuit, while the combination of each with the carrier applied at C is transmitted. Thus if the side-band and received carrier enter together at S, the components having the original signal frequencies are eliminated and only the displaced components remain.

When the other side-band is added, the situation is still further complicated. In the absence of received carrier, the local carrier and one side-band give a set of components the frequencies of which are greater than those of the signal by n, while the carrier and other side-band give a set less by the same amount. These two sets combine in much the same way as do the displaced and normal speech obtained with a single side-band and received carrier. Here, however, the beat frequency is $2\,n$. Also, as the two sets are equal in amplitude, the beats will be much more pronounced, the intensity falling to zero each time the two waves are in opposition. For slow beats the

apparent pitch of the sound is half way between the frequencies of the two equal components, and so the normal voice frequency will be heard. With large frequency displacements of the carrier the two displaced speech waves, being of equal intensity, will be more easily distinguished than in the case of the single side-band.

It is interesting to note that this result, as well as the frequency shift that occurs with a single side-band, follows directly from the relations arrived at above for the phase displacement of a synchronous carrier. The non-synchronous carrier may be thought of as a synchronous one the phase of which is varied with the frequency of the departure from synchronism. With a single side-band it was shown that a phase displacement of the carrier affects only the phase of the reproduced component and that it changes this by an amount equal to its own displacement. This progressive phase displacement in all the components of the reproduced wave is, of course, equivalent to a change in their frequencies equal to the frequency displacement of the carrier. With both side-bands present, a phase displacement was shown to have no effect on the phases but to change the amplitudes of the reproduced components by the factor $\cos \eta$. Thus a progressive change in η will cause a cyclic variation in amplitude having two minima for each cycle of η; that is, a frequency of $2n$.

If the two side-bands are accompanied by the carrier there are added the beat note of the two carriers and the components of the original signal. The addition of a small amount of this speech of uniform amplitude to that of varying amplitude already present merely tends to make the variation slightly less pronounced. From the foregoing it appears that for telephony the most favorable condition for using a local carrier which is out of synchronism is that in which only one side-band is transmitted. Fairly considerable frequency variations are then permissible and asynchronous operation appears to have practical possibilities.

For telegraphy the case is quite different. In the first place the important components of telegraph signals are much lower in frequency, so that the side-band lies closer to the carrier, and a much smaller absolute displacement of the carrier frequency is needed to give the same effect as in the telephone case. Considering only such small displacements, it appears that the general addition or subtraction of frequencies which occurs with a single side-band will alter the shape of the signals quite seriously. The slow fluctuations in signal intensity which occur with both side-bands are probably less serious over most of the cycle. However, they might well cause some signals to be lost entirely each time the intensity passed thru zero.

For asynchronous reception, then, just as for the case of a local carrier displaced in phase, single side-band transmission is preferable for telephony and double side-band for telegraphy. The difficulties are of the same general nature in the two cases, but with the asynchronous carrier they are considerably greater. This is in agreement with the idea expressed above, that lack of synchronism may be looked upon as an aggravated case of phase displacement.

Public Address Systems[1]

By I. W. GREEN and J. P. MAXFIELD

SYNOPSIS: A public address system comprises electrical equipment to greatly amplify a speaker's voice so it will reach a much larger assemblage than he could speak to unaided. Beginning with the presidential conventions of the two major parties in 1920 and the inaugural address of President Harding in March 1921, when a special address system installed by the telephone engineers enabled him to address an audience estimated at 125,000, there followed in rapid succession, many public events demonstrating the value of such systems. One of the most notable of these occurred on Armistice Day 1921, when the speeches, prayers and music at Arlington, Virginia, were heard, not only by 100,000 persons gathered there at the National Cemetery, but by some 35,000 in New York City and 20,000 in San Francisco. On this occasion the three public address systems, one for each of these cities, were joined by long distance telephone circuits.

The fundamental requirements of a satisfactory public address system are naturalness of reproduction and wide range of output volume. The meeting of these two requirements for music proves more difficult than for speech.

The public address system here described is most readily considered in three sections—"pick-up" apparatus which is placed in the neighborhood of the speaker and converts his words into undulatory electric currents; a vacuum tube amplifier for amplifying these currents; and a "receiver-projector" for reconverting the current into sound waves and distributing the sound over all of the audience. In the present system each of these three parts of the equipment has been designed with the intention of making it as nearly distortionless as possible, so that the various parts might be adaptable for audiences ranging in size from possibly one thousand to several hundred thousand, and might also be used in connection with the long distance telephone lines and with either radio broadcasting or receiving stations. One of the larger public address systems is easily capable of magnifying a speaker's voice as many as 10,000 times.

The pick-up device whether of the carbon microphone variety or a condenser transmitter need not be placed close to the speaker's lips but will operate satisfactorily when four or five feet away. The loud-speaking receiver mechanism is so designed that it will carry a power of several watts with small distortion. Under normal conditions, 40 watts distributed among a number of receiver-projectors arranged in a circle is ample to reach an audience of 700,000 persons.—*Editor.*

THIS paper aims to present the problems encountered in the development of electrical systems for amplifying the voices of public speakers and music; and to describe the equipment as brought to a commercial state and now in use in the United States and various other countries.

The two main requirements of a successful public address or loud speaking system are, first, that it shall reproduce the sounds, such as speech or music, faithfully; and second, that this faithful reproduction shall be loud enough and sufficiently well distributed for all of the audience to hear it comfortably. Most of the development work

[1] Presented at the Midwinter Convention of the A. I. E. E., New York, N. Y., February 14–17, 1923. Published in the Journal of the A. I. E. E. for April, 1923.

has been directed toward obtaining these two results under the various conditions which surround the operation of these systems.

The faithful and natural reproduction of sound depends upon many factors, of which the following are some of the more important: The acoustics of the space in which the sound originates, the characteristic of the loud speaking system itself and the acoustics of the space in which the sound is reproduced. Where the sound is picked up and reproduced in the same space, as is the case when the speaker is using one of these systems to address a large audience locally, there is a reaction between the horns and the transmitter or pick-up mechanism which is controlled by the acoustic conditions under which the system is operated.

Acoustics of the Space

In connection with the acoustics of the space in which the sound originates, or in which it is reproduced, four factors stand out. These concern the effects of reverberation, of echo, of resonance, and of diffraction. In the specific cases where the sound is reproduced in the same space or room in which it originates, another effect is encountered, which has generally been termed "singing," and is evidenced if sufficiently great by the emission of a continuous note from the equipment.

Reverberation is caused by reflection and is evidenced by the persistence of the sound after its source has ceased emitting. When the reverberation in the space in which the initial sound is being picked up is sufficient to cause one sound to hang over and become mingled with succeeding sounds, in other words, so that the sound from one syllable interferes with that of the succeeding syllable, it is practically impossible to improve the acoustic conditions solely by the use of the public address system. In such a case, the first procedure is to place material which absorbs sound in the space. The purpose of this absorbing material is to lower the time required for the sound to die away after the source ceases to emit. The amount which any given material lowers the time of reverberation depends not only upon the amount of material introduced, but also upon its disposition within the space.

The term "echo" applies to a similar phenomenon, but is generally used where there is sufficient time lag between the reflected sound and that originally emitted, so that two distinct impressions reach the ear.[2]

[2]Collected Papers on Acoustics, Wallace Clement Sabine, Harvard University Press, 1922.

The troubles encountered from echoes usually occur only in large buildings or large open spaces surrounded by buildings, trees, or other obstacles and are generally associated with interferences with the reproduced sound rather than with the original sound. There are cases, however, particularly in auditoriums, where some of the walls or ceiling are large curved surfaces, in which case localized echoes may result. The speaker's voice or extraneous sounds from the audience may be reflected from one or more of these surfaces to focus spots where the volume of sound is consequently abnormally great. It is important, therefore, that the transmitter which is picking up the sound shall not be located at one of these spots. These points of localized echo are particularly troublesome also when they occur in the space occupied by the audience. Under these conditions not only is the sound intensity too great, but the character of the sound is altered and very often badly confused. The avoidance of such difficulties is a matter of test and the proper arrangement of the reproducing mechanism, as will be seen later in some detail.

The effect of resonance seldom occurs in connection with the amplified and reproduced sound, inasmuch as the spaces dealt with are large and their natural frequencies are too low to be troublesome. Resonance usually becomes of importance in connection with mounting the pick-up apparatus or transmitter. It generally results from attempts to conceal the transmitter by placing it in some form of small enclosure. The best form of housing from an acoustic standpoint consists of a screen cover which protects the instrument from being struck or injured but in no way affects the sound reaching it.

Resonance produces a distortion which it has been customary to consider as of two varieties. First, there is an unequal amplification of sounds of various frequencies and second, there is the introduction of transients. These transients occur whenever the sound changes but are most easily recognized audibly by their continuation after the source has ceased emitting. They also have frequency characteristics which depend not only on the sound which started them but also upon the character of the resonant portion of the system.

The troubles introduced by diffraction are seldom of very great importance except where the sound is reflected from regularly spaced reflectors or passed through regularly spaced openings. Quite serious diffraction troubles have been encountered when operating a loud speaker in a large field, surrounded by an open work board fence, the trouble being evidenced by very distinct areas, particularly at the outskirts of the audience, where the sounds were badly distorted.

The difficulties encountered as a result of "singing" form one of the most troublesome problems connected with the actual operation of these systems. When a portion of the sound emitted by the projectors reaches the transmitter with sufficient intensity, that its reproduction is as great as the originally emitted sound from the projectors, and with such phase relation that it tends to aid the original sound, the system will emit a continuous note. Moreover, when the portion of the sound from the projectors which reaches the transmitter is not sufficient to cause a continuous note, it may be sufficient to cause considerable distortion of the speech or music. In excessively reverberant halls these conditions are often fulfilled when the actual amplification is so small that the people at the distant points are scarcely able to hear the speaker. In all cases in our experience the difficulty has been sufficiently overcome by properly placing the transmitter with respect to the projectors. The situation is very much helped by the presence of the audience, which adds considerably to the acoustic damping of the room.

It will be seen, therefore, that the acoustic conditions of the space in which loud speakers are used are of considerable importance.

Characteristics of the System

The first requirement of the system itself is that it shall reproduce speech or music faithfully. A system is said to do this, or in other words, its quality is called perfect, when the reproduced sound contains all of the frequencies, but no others, contained in the original sound striking the "pick-up" mechanism, and when these frequencies have the same relative intensities that they had in the original.

An imperfect or distorting system is one which fails to fulfill this requirement. There are two main types of distortion which had to be considered; the first being the unequal amplification of the system for the various frequencies constituting the sound and the second being the introduction of frequencies not present in the original sound. For simplicity of discussion, this last class will be divided in three parts, namely: the effect of transients, the effect of asymmetric distortion and the effect of disturbing noises.

The effect of transients has already been mentioned in connection with acoustics and they, of course, produce the same type of distortion whether they occur in the acoustic or the electrical system. Transients occur whenever the sound changes either in pitch or intensity, and are introduced at the beginning and ending of each speech sound. This modification of the characteristics of the speech

sounds acts to lower the intelligibility. It probably causes more trouble in speech transmission than the fact that the sound continues after the source ceases.

Asymmetric distortion affects one half of the wave differently from its other half. This causes the introduction of frequencies which, in some cases, produce very serious disturbances in the transmission of music and speech. The most noticeable troubles are from the formation of sum and difference tones.[3] Such tones are likely to give rise to dissonances with the other sounds occurring in the music. In the case of speech asymmetric distortion manifests itself by a lower intelligibility.

The effect of foreign noises sometimes encountered is twofold. First, they influence the ability of the listener to hear the characteristics of the speech sounds and hence tend to lower the intelligibility. Secondly, the constant attempt of the hearer to sort out the speech sounds or music through the disturbing noises tires him appreciably. In order that this strain shall be inappreciable, it is desirable that the sound delivered by the system shall be at a power level approximately 10,000 times that of the noise.

The second general requirement which is placed on a successful system is that it shall deliver its faithful reproduction loud enough for all the audience to hear it comfortably and enough above noise for good intelligibility. In this connection there have arisen one or two interesting points bearing on the psychology of hearing. One of the most striking of these is concerned with the coordination between hearing and seeing. Although the projectors are usually mounted twenty or more feet above the speaker's head, and in some exceptional cases, slightly to one side of him, the majority of the audience is conscious of only one source of sound, and that appears to be the speaker himself.

This phenomenon is so marked that in several cases the question has been rasied in the minds of the listeners as to whether the system was functioning. They could only be convinced that it was by having it shut down for a few seconds when their inability to hear made them realize how successfully the system could operate.

Another of these psychological phenomena deals with the apparent distortion of the voice when its intensity at the ears of the listener is too great or too small. If the speaker is talking in a normal conversational tone, his voice contains a larger percentage of low frequencies than is the case when he is raising his voice to a considerable

[3] Origin of Combination Tones in Microphone-Telephone Circuits. E. Waetzmann, Annalen der Physik, Vol. 42, 1913.

volume. If the loud speaker so amplifies this voice that it reaches the audience with such volume that their instinct tells them that the speaker should be shouting, the system appears to make his voice sound quite heavy and somewhat unnatural. It has been found necessary, therefore, to so regulate the amount of amplification that the people at the furthest portion of the hall can hear comfortably and the volume of sound shall not be permitted to become any louder than necessary to meet this condition. On the other hand, if the volume is insufficiently loud, certain of the weaker speech sounds are entirely lost, and it becomes difficult to understand.[4]

Solution of the Problem

With these considerations in mind it may be interesting to take a brief survey of the whole problem and the method by which the solution was reached. Two general methods of attack were considered. The first was to attempt to make each unit of the system faithfully reproduce its input, while the second was to make any distortions of one part of the system, cancel those of another portion, so that the complete system would operate satisfactorily. In either case, it was desirable to keep each unit free from asymmetric distortion, as this type of distortion cannot be easily compensated for.

While it would probably have been simpler to follow the second line of attack, the greatly increased flexibility of a system in which each part is correct in itself was of sufficient value to cause the attempt to be made that way. When it is realized that these systems, to be commercially successful, must be capable of operating for various sized audiences, ranging possibly from one thousand to several hundred thousand, that they must be used in connection with long distance telephone lines, as well as with either radio broadcasting or receiving stations, the desire for flexibility can be understood.[5]

As a result of attempting the development in the manner already described, there has resulted a system which involves four functional units; a "pick-up" mechanism or transmitter unit, a preliminary amplifier unit, commonly called the speech input equipment, a second amplifier unit commonly called the power amplifier, and a receiver-projector unit for transforming the amplified currents back into sound, and properly distributing it throughout the space to be covered.

[4] Physical Examination of Hearing, R. L. Wegel, *Proceedings of the National Academy of Sciences*, Volume 8, Number 7, July, 1922.

[5] Use of Public Address Systems with Telephone Lines, W. H. Martin and A. B. Clark. Presented before A. I. E. E., Feb. 14, 1923.

It may be interesting at this place to determine how successfully these various units and the system as a whole fulfill the requirements of equal sensitivity to all frequencies within the important speech or music range. Fig. 1 shows the relative sensitiveness of the trans-

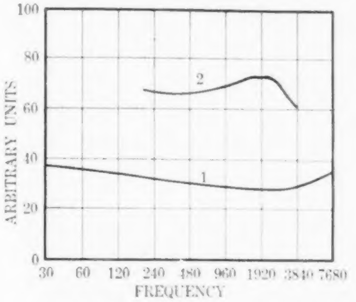


Fig. 1—1-Condenser Transmitter with Associated Amplifier.
2-Carbon Transmitter without Amplifier.

mitter as a function of frequency. The ordinates are proportional to the logarithm of the power delivered for constant sound pressure at the diaphragm and the abscissae to the logarithm of the frequencies employed. The lower of the two figures refers to the condenser transmitter with its associated input amplifier.[6] The upper refers to the push-pull carbon-type transmitter. These transmitters will be described in detail later.

Fig. 2 shows a similar curve for the complete amplifier system, comprising a three-stage speech input amplifier, and a power stage capable of delivering approximately 40 watts of speech frequency

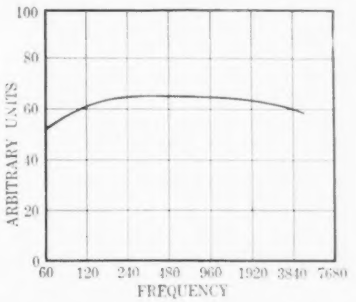


Fig. 2—Public Address System Amplifiers.

[6] The Sensitivity and Precision of the Electrostatic Transmitter for Measurement of Sound Intensities. E. C. Wente, *Physical Review, N. S.* Vol. 19, No. 5, May, 1922.

electrical power without distortion. In connection with these amplifiers a sharp distinction should be made between their gain rating, or amplification, and their overload or power rating. Gain measures the power amplification which can be obtained provided the input is small enough so that the equipment at the output end is not overloaded. Overload or power rating refers to the maximum power which can be supplied by the amplifier without causing distortion of the currents being amplified. Although the power rating of power equipment is usually determined by the heat which can be dissipated, a marked distortion of wave form takes place when the iron in any of the apparatus is worked beyond the straight line portion of the magnetization curve. In the case of amplifiers, the maximum power obtainable is limited by the power output at which distortion occurs rather than by the heat which can be dissipated.

Fig. 3 shows a chart for the characteristics of the complete system,

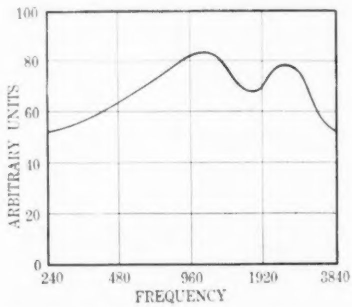


Fig. 3—Complete Public Address System with Carbon Transmitter.

including the carbon transmitter, the speech input and power amplifiers and the receiver projector unit.

In connection with the requirements for equal amplification of all frequencies, it is interesting to note that a system, which does not fail to reproduce equally all frequencies in the speech range by more than a ratio of 10 to 1 in reproduced power, is indistinguishable from a perfect system or from the speaker, himself. It seems probable that this effect is, in some way, connected with variations in the frequency sensitiveness curve of normal ears. Normal ears show a sensitiveness variation with frequency as great as 10 to 1 and the frequencies of maximum sensitiveness vary materially from one individual to another.[7]

[7] Frequency Sensitiveness of Normal Ears, by H. Fletcher and R. L. Wegel, *Physical Review*, July, 1922.

It has been found that in order to transmit speech with entire satisfaction for loud speaker purposes, that is, sufficiently well so that the audience is not aware of the contribution of the mechanical equipment, it is necessary for the system to operate with essentially uniform amplification over a range of frequencies from 200 to 4000 cycles. While there are, in speech, frequencies slightly outside of this range, the loss in naturalness and intelligibility by the system's failure to reproduce them, is slight.[8]

While no such frequency range is required for intelligibility only, it has been found that systems not covering substantially this frequency range, sound unnatural. When the lower frequencies are missing, the reproduction sounds "tinny." When the higher frequencies are missing it sounds heavy and muffled. The requirements for thoroughly natural reproduction of music are probably more severe, particularly in the low-frequency region, than are the similar requirements for speech, but, at the present time, complete data are not available to indicate the contribution of these frequencies to naturalness.

In connection with the flexibility of the system, it is interesting to note that the speech input equipment has been designed to raise the power delivered by the transmitter to such an extent that it is sufficient for long distance telephone transmission or for the operation of a radio transmitting set. The power amplifier is designed to receive power at approximately this level and deliver it to the projector units sufficiently amplified to operate them satisfactorily.

Future Development

In viewing the loud speaker field from the point of view of future development there are two lines of attack along which work is being done, and which give promise of success. These are the improvements in frequency characteristics and increase in the range of loudness which the system can accommodate satisfactorily.

The improvement to be expected from a more uniform frequency characteristic is mainly an increase in naturalness, especially where music is being reproduced. A slight increase in intelligibility may be hoped for, although this factor is of little importance, as the present system is satisfactory in this respect.

The other improvement mentioned, namely, the volume range, is probably the more difficult, but is necessary before music can be re-

[8] The Nature of Speech and Its Interpretation. Harvey Fletcher, *Journal of the Franklin Institute*, Vol. 193, No. 6, June, 1922.

produced in a perfect manner. Rough experimental data indicate that the loudness in an orchestra selection may vary from one part of the selection to another by a ratio as great as 50,000 to 1. While the present equipment does not operate with entire satisfaction over this range of loudness, it has been found relatively easy to obtain good results by manual adjustment of the amplification during the rendering of the selection. If the gain is varied in small enough steps, the change is not noticeable to the listeners.

An increase in the loudness range would render the manual adjustment unnecessary and would make the reproduction a faithful duplicate of the music as actually played.

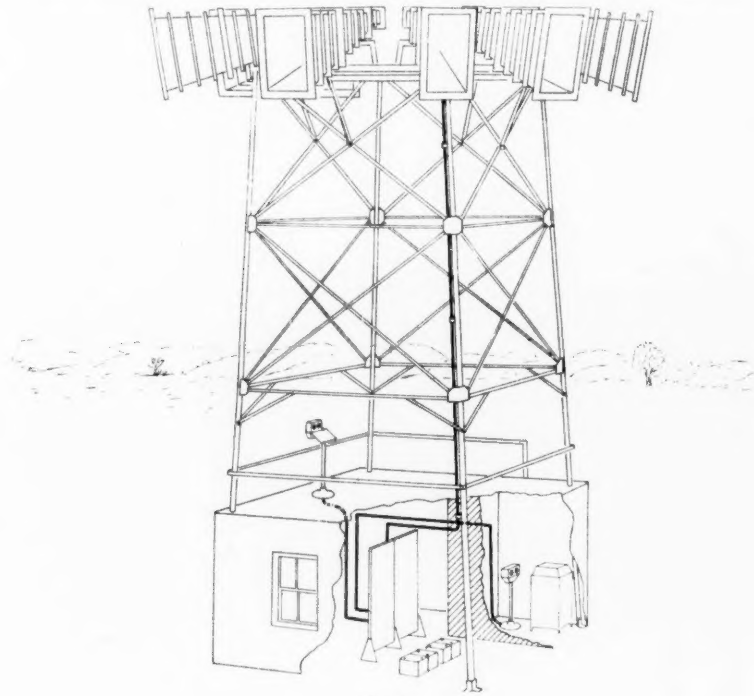


Fig. 4.

Technical Description of the System

The foregoing discussion having described the requirements which must be met in order that the public address system shall successfully transmit speech and music, the system in its commercial form will

now be described. In order to make clear the arrangement of the equipment, a typical installation is shown in Fig. 4, this being an installation where the audience and speaker are in the open air, and where no connection is made with the long distance lines. It might be well to state here that with the equipment shown an audience of 700,000 can be adequately covered.

Some of the sound leaving the speaker's mouth is picked up by the transmitter, on a reading-desk type of pedestal, which is normally mounted at the front of the platform.

The feeble currents from the transmitter are led by carefully shielded leads to the amplifiers in the control room, which is usually

Fig. 5.

located directly beneath or to one side of the speaker's stand. A floor space of not more than 125 square feet is required for this room, even in the case where it is desirable to transmit phonograph music to the audience between speeches.

The amplifier and power supply equipment is shown on the two panels in the center of the control room. The amplified speech currents are led from these amplifiers to the receiver projector units, which, in this case, are arranged on the super-structure above the

speaker's platform. This position is most desirable, as the illusion produced is such that the voice appears to come from the speaker rather than from the projectors, a factor, the importance of which has already been mentioned. Moreover in this position the acoustic coupling between the transmitter and the projectors is a minimum, permitting the operation of the system at a satisfactory degree of amplification with an ample margin below the point where singing troubles are encountered.

A public address equipment, similar to that just described, but with a somewhat lower power output, has been developed for use at the smaller open air meetings, and in all but the largest indoor auditoriums. Fig. No. 5 shows one of these equipments, mounted on an automobile truck, which has been employed at a number of points in the eastern part of the United States. This smaller system has characteristics as good as the larger system in regard to faithful reproduction of speech and music, with a power output in the order of one-tenth as great. An audience of 50,000 can be adequately covered at an outdoor meeting with this system.

Fig. No. 6 is a schematic arrangement of the equipment at an installation of the type shown in Fig. 4. At the extreme left are the

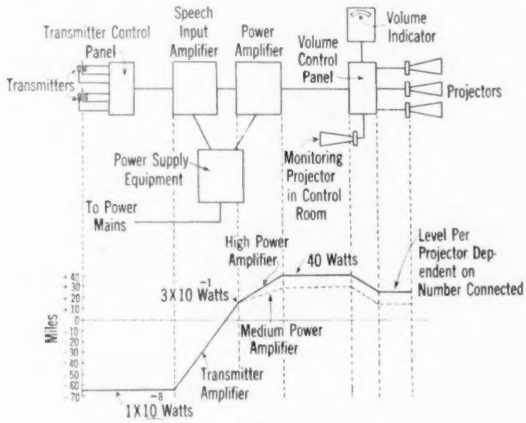


Fig. 6.

transmitters where the sound waves are picked up. The output from these transmitters is taken to a switching panel where means are provided for cutting in the various transmitters. From this panel the transmitter currents are taken to the transmitter amplifier, which is capable of amplifying them to a power level suitable for input to

the power amplifier, or for connection to the long distance lines, in those cases where the speeches are transmitted to distant audiences. It is also suitable where connection is to be made to a radio station for broadcasting the speeches. The power amplifier is shown just to the right of the transmitter amplifier. Below it is indicated the power supply equipment by which the commercial power is converted to a form suitable for the vacuum tubes in both amplifiers. The output from the power amplifier is taken through a panel where switches and a multi-step auto-transformer are provided for the regulation of several projector circuits. Just above this panel is an indicating instrument, known as the volume indicator, provided in order that the operator may know what volume output is being delivered to the projectors.

The projectors, at the extreme right of the figure, consist of the motor or receiver unit transforming the speech currents into sound waves, and a horn to provide the proper distribution of the sound.

It is interesting to note the power amplification which is obtained in the larger of the two systems from the transmitter to the projectors. Referring to Fig. No. 6, a chart will be seen which indicates the power levels through the system drawn to a scale based on miles of standard telephone cable, our usual reference unit. The output of the high quality transmitter is of the order of 65 miles below zero level, this latter being the output from a standard telephone set connected to a common battery central office by a line of zero resistance. Expressed in watts the output of this transmitter under average conditions of use with the public address system is of the order of 10^{-8} watts. Incidentally, this is of the same order as the speech power picked up by the transmitter, or in other words, the transmitter does not amplify the speech power received by it as is the case with the transmitter used for regular telephone service.

This very minute amount of power in passing through the transmitter amplifier may be amplified about 120,000,000 times. Expressed in terms of telephone power levels, this is 17 miles above the zero level previously mentioned, or a few tenths of a watt.

The power amplifier serves to increase this power from a level of 17 miles to about 40 miles, the latter corresponding to about 40 watts. This power is then distributed to the projectors, the amount consumed by each projector, of course, depending upon the number connected.

An idea of what this amount of power at speech frequencies means may be given by the statement that it is sufficient to operate at about the level considered commercial all of the 14,000,000 tele-

phone receivers in use in the Bell system if these were directly connected to the amplifier.

In describing the various pieces of equipment which together make up the system, we will follow the order in which the power is carried through the system from the transmitter to the receiver-projector units where the amplified sound waves are propagated.

Transmitters

In the early work on the public address system, an air-damped, stretched diaphragm condenser transmitter was employed, having a thin steel diaphragm about 2 inches in diameter, constituting one plate of the condenser. The other plate was a rigid disk, the dielectric being an air film 1/1000 of an inch in thickness. Due to the stretching of the diaphragm and the stiffness of the air film, the diaphragm of this transmitter had a natural period of approximately 8000 cycles per second which is well above the important frequencies in the voice range. This high natural period, in conjunction with the damping due to the thin film of air, resulted in a transmitter of very high quality of reproduction. However, its extremely small capacity (in the order of 400 micro-microfarads) made it necessary to use leads of very low capacity between the transmitter and the first amplifier, and due to the high impedance of the transmitter and its associated input circuit to voice frequency currents, these leads were very susceptible to electrostatic and electromagnetic induction. It was necessary to limit them to a length of 25 feet, and to provide complete shielding. Moreover the output of this transmitter was less than one five-thousandth of that of the transmitter now used, and for the early installations of the system, it was necessary to provide a preliminary amplifier beneath or to one side of the speaker's stand in order to keep the transmitter leads short and to provide sufficient power to properly operate the main amplifiers. Work was therefore undertaken to provide a transmitter having quality practically as good as the condenser transmitter, volume output sufficient to operate the main amplifiers, and not requiring the elaborate precautions as to shielding the leads.

The high quality transmitter which was the result of this development work is of the granular carbon type with two variable resistance elements, one on each side of the diaphragm and is commonly known as a push-pull transmitter. It has nearly the same high quality reproduction characteristics as the condenser transmitter, due to the use of the same stretched diaphragm and air damping structure. It

introduces no appreciable distortion over the range of frequencies required for good reproduction of speech, but it must be understood, that this quality was obtained only at the sacrifice of sensitiveness, the latter being in the order of 1/1000th that of the transmitter used at telephone stations in the Bell system. With the multi-stage vacuum tube amplifiers available this low volume efficiency is not serious, and in fact we are using this transmitter for what is known as distant talking, *i.e.*, the speaker may be at a distance of five or six feet from the transmitter. This is, of course, necessary in any transmitter suitable for public address work as it is not possible to greatly limit the movement of the speakers, nor can they be required to use a hand transmitter. It might be well to point out that this sacrifice in volume efficiency in order to gain high quality is possible at the transmitting end of the system, but not at the opposite end where the electrical power is transformed into sound waves and propagated, as the device at this point must be capable of handling large amounts of power with minimum distortion.

Referring to Fig. No. 7 which is a cut-away view of this push-pull high quality carbon transmitter, the granular carbon chambers will be seen. The electrical path through each of these variable resistance elements is from the rear carbon electrode through the carbon granules

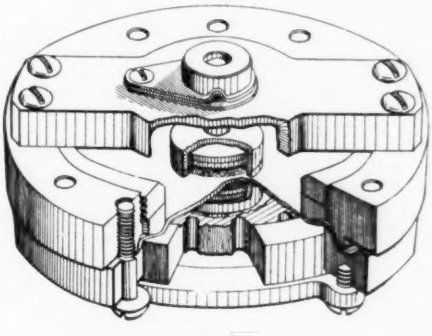


Fig. 7.

to a gold-plated area on the diaphragm itself. The resistance of this path is about 100 ohms and as the two buttons are in series for the telephone currents, the transmitter is designed to work into an impedance of 200 ohms. The double button construction almost completely eliminates the distortion caused by the non-linear nature of the pressure-resistance characteristics of granular carbon.

As this instrument has a practically flat frequency characteristic no collecting horn or mouth piece is used with it as resonance is introduced by such chambers, with accompanying distortion. To insure the insulation of the transmitter from building vibrations, a simple spring suspension has been provided. To protect the transmitter from injury, two types of transmitter mountings have been used, both arranged for the suspension of the transmitter in a screen-enclosed space—the first adapted to take a single transmitter for indoor use only, while the second for outdoor use, mounts two transmitters within a double screen enclosure to prevent any noise effects due to winds. This second type is arranged to attach to a simple pedestal-type of reading desk, which it has been found desirable to provide as there is a slight tendency for the speaker to remain fairly close to the desk. In this connection it is interesting to note that we have found a small rug, so placed as to cover the area which the speaker should occupy during the delivery of his speech, is of great assistance in this regard, as he unconsciously confines himself to the area of this rug. Both of these measures to insure the speaker remaining in proper relation to the transmitter, are supplemented, wherever possible, by an explanation of the system to all the speakers previous to the actual performance.

Transmitter Switching Panel

Resuming the path of the speech currents through the system, the output from the transmitters is taken to a panel designed to enable the operator to switch quickly from one transmitter to another, as with some public functions, the speeches are made at different points during the ceremonies. This switch is arranged to short-circuit the output of the power amplifier when passing from one transmitter to another, to prevent clicks in the projectors. In certain cases, the equipment is arranged to permit two or more transmitters to be connected to the amplifiers at one time, as is desirable when solo singers and an orchestra are to be picked up in a theatre, with proper adjustment of their respective volumes.

The amplifier equipment has been built in four units which may be grouped as necessary under the various conditions encountered in commercial installations. The proper amplifiers are determined, first, by the source of the voice frequency current to be amplified, that is whether a distance talking or a close talking transmitter is to be used, or whether the speeches are brought in over a telephone line, and secondly, the size of the space in which the amplified sounds

are to be delivered to the audience. It was found that four units would provide for all the conditions occurring in practise, two of these being speech input or transmitter amplifiers with different gains and two being power amplifiers of different power ratings. These units and other equipment used with the system, are made up in panels, of uniform width, in order that the proper equipment for any installation may be assembled on two vertical angle iron racks arranged to be fastened to the control room floor.

SPEECH INPUT AMPLIFIER—FIRST TYPE

The first of the speech input amplifiers is shown schematically in the upper part of Fig. 8. It is a three-stage amplifier. Two po-

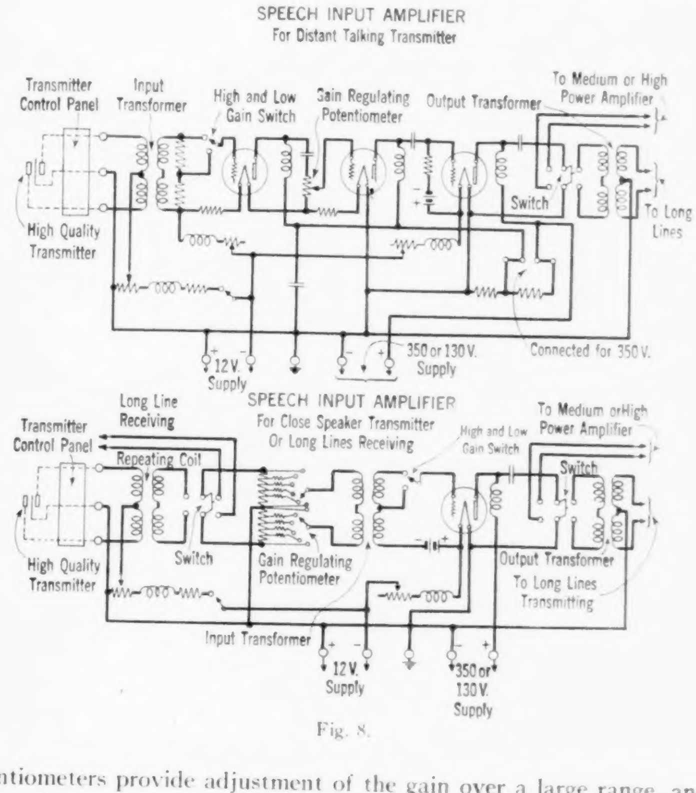


Fig. 8.

tentiometers provide adjustment of the gain over a large range, and switching arrangements allow the output to be connected directly

to the input of the power amplifier, when the program is to be transmitted to a local audience; or to be connected, through a transformer of proper impedance, to the long distance lines when the program is to be transmitted to a distant audience, or to a radio-broadcasting station. The filaments of the tubes are supplied from a 12-volt storage battery, while the plate circuits obtain direct current at 350 volts from the power supply equipment mentioned later. Arrangements are also provided for using 130 volts instead of 350 volts under certain conditions. The proper grid potentials are obtained by utilizing the drop over a resistance in the filament circuits of the first two tubes, and for the third tube small dry cells furnish the grid potential. The maximum gain with this amplifier is 85 miles, which expressed as a power ratio is 1.2×10^8. Under this condition the output is approximately 3/10 of a watt. The front and rear views of this amplifier, mounted on the supporting rack, as shown in Fig. 13, where the gain regulating potentiometer, the rheostats for controlling the filament and transmitter currents, the three tube mountings with protective gratings and the jacks which permit the connection of instruments for determining the current flow in the filament, plate and transmitter circuits, will be noted. Great care was taken in the design of this amplifier to obtain as nearly as possible equal amplification of all the important frequencies in the voice range. The transformers, and the retardation coils in the plate circuits were chosen with this consideration in mind.

Power Amplifiers

For practically all of the larger installations the maximum power possible with the system is required and the output from the transmitter-amplifier is taken directly to the high power amplifier. Referring to Fig. 9 it will be seen that this is a four-tube amplifier so connected that but one stage of amplification is obtained. Usually alternating current at 12 to 14 volts is used for heating the filaments of these tubes, the latter being connected in what we know as a push-pull arrangement. It will be seen that each side of the push-pull arrangement consists of two power tubes in multiple. It is interesting to note that this push-pull arrangement of the tubes will deliver somewhat more power for equal quality than the same number of tubes connected in the ordinary multiple arrangement, since the tubes may be worked beyond the straight part of their characteristic. The grid potential is chosen to permit the largest variation of current without distortion and is obtained from a group of small flashlight batteries.

The output transformer at the right of the figure is designed to match accurately the impedance of the tubes to that of the number of receiver-projector units which has been found to give the greatest flexibility under the varying conditions of commercial operation. This amplifier, speaking in telephone terms, is worked at a gain of 23 miles, a power amplification ratio of about 200, the

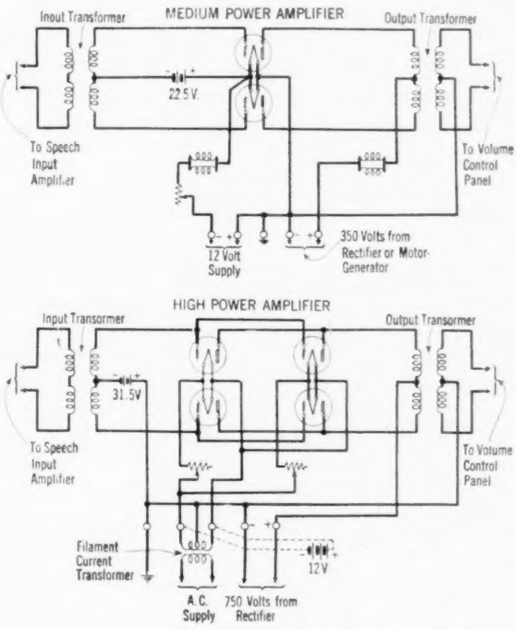


Fig. 9.

maximum output being about 40 watts. The plate circuits of the tubes are supplied at a d-c. potential of 750 volts. As has been pointed out previously, this amplifier gives a practically uniform gain for all the important frequencies in the voice range. This high-power amplifier is shown mounted on the supporting rack, in Fig. 13. The apparatus on the rear of the panel is protected with a sheet metal cover and integral with this cover is a disconnecting switch, which, when the cover is removed, cuts off the high potential from all the exposed parts on the set.

For indoor installations where the audience is small the power output given by the high-power amplifier is not required and a medium-

power amplifier has been developed for this use. It is arranged to connect directly to the transmitter amplifier and the output is taken to the projectors through the volume control panel. It has a gain of 17 miles or a power amplification ratio of about 33. The maximum output is about 4 watts, or about one-tenth of the power obtainable from the high-power amplifier.

The schematic of this amplifier, is shown in Fig. 9. The input coil is the same as is used in the high-power amplifier. The push-pull connection of the tubes is also used in this amplifier, although but two power tubes are used. The filaments of these tubes are supplied from a 12-volt storage battery while the plate circuits are supplied at 350 volts direct current from a motor generator set which will be described later.

Speech Input Amplifier—Second Type

A speaker using the system may read his speech from his home or office and in such cases it is unnecessary to use the push-pull carbon transmitter in the distant-talking manner. For use when this transmitter is spoken into from a distance of a few inches, a second form of speech input amplifier has been made available having a gain of the proper value to supply either of the power amplifiers, or a long distance line if desired. This gain is relatively small as the output of the transmitter when used for close talking is about 10,000 times that when it is used for distant talking.

Fig. 8 shows the schematic of this amplifier which is a single-stage one, employing one tube and having the same over-load characteristic as the first form of speech-input amplifier. A two-way switch permits the connection of the transmitter or an incoming long distance line to the amplifier. To the right of this switch is a potentiometer for regulation of the gain. To the right of the tube is a second two-way switch for connecting the output either to the power amplifier or through an output transformer to an outgoing long distance line. The power supply for the tubes and transmitters, is the same as was described under the first form of speech input amplifier.

The switching means provided on this amplifier allow it to be used in a number of ways. Announcements from a close talking transmitter may be made from the projectors through a power amplifier or may be sent out on the telephone lines to a distant public address system installation or a radio-broadcasting station. In addition to these uses, incoming speech over the long distance lines may be put out on the projectors through the power amplifier or may be sent out on the long distance lines to a distant installation.

Volume Control

As discussed heretofore, it is necessary to give the operator control of the volume put out by each projector or group of projectors. The equipment provided for this purpose is mounted on a panel uniform with the others and consists essentially of an auto-transformer connected across the output of the power amplifier with 11 taps multipled to the contacts of eight dial switches, the arrangement being shown schematically in Fig. 10. Seven of the dials control projector circuits on each of which one or more projectors may be grouped, the

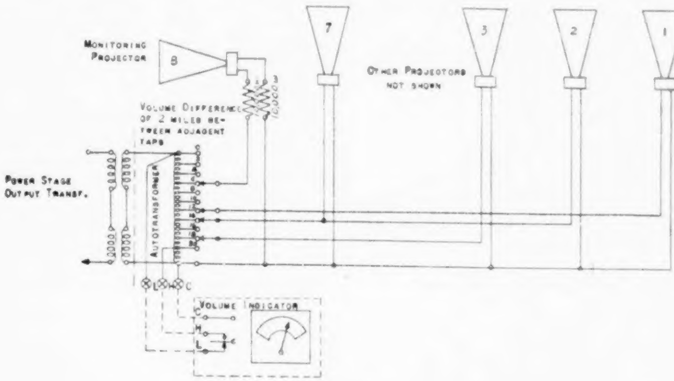


Fig. 10.

eighth dial being reserved for controlling the volume of the operator's monitoring projector in the control room. A key is associated with each dial for opening the circuit and a master key is provided for cutting off all of the projectors simultaneously.

The device shown as "Volume Indicator" in this figure consists of a vacuum tube detector bridged across the output terminals of the power amplifier. The rectified current is taken to a sensitive direct-current meter of the moving coil type, the degree of deflection of this meter measuring the output from the power amplifier when connected at a proper place in the circuit. The deflections of the meter therefore serve as a basis for determining the adjustment required on the transmitter-amplifier to give the required output when switching from one transmitter to another or for different speakers.

Receiver—Projectors

From the control panel the power is taken to the projectors, each of these consisting of a loud-speaking receiver mechanism to transform

the speech-currents into sound waves, and a horn to distribute the sound. The receiver is so designed that it will carry several watts

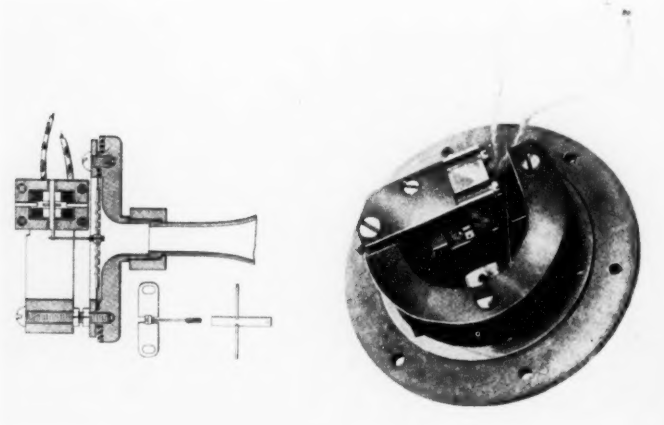


Fig. 11.

with small distortion. It is shown in Fig. 11 where it will be seen that a light spring-supported iron armature is mounted between the poles of a permanent magnet and passes through the center of the

Fig. 12.

coils carrying the voice currents. A light connecting link ties one end of this armature to the diaphragm which is of impregnated cloth, corrugated to permit vibrations of large amplitude. A stamped metal

cover protects the parts from mechanical injury, and a cast iron case, in which the whole assembly mounts, is provided for protection against moisture.

One of these receivers equipped with the largest projector provided, will carry without serious distortion or overheating, power which is

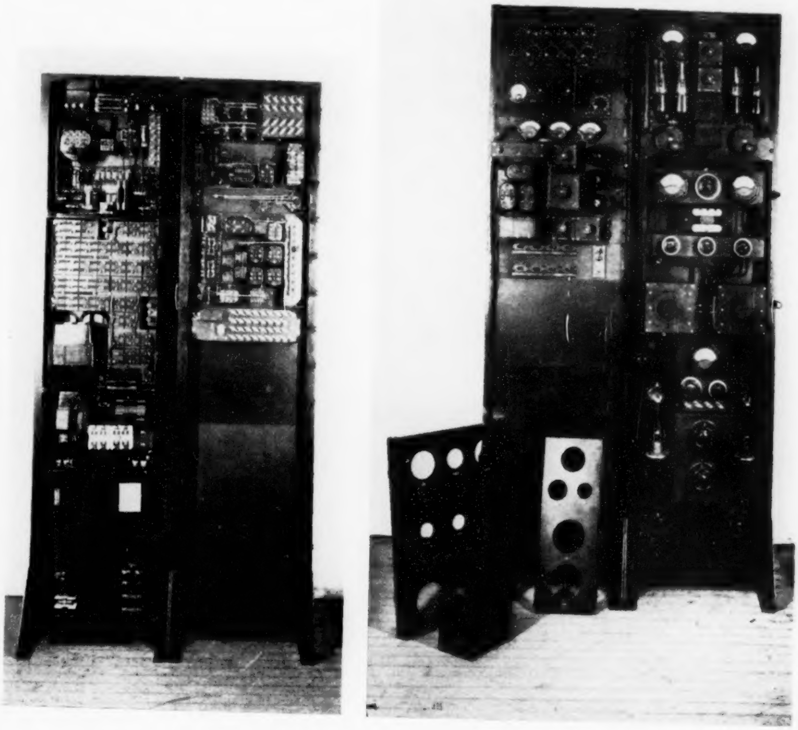


Fig. 13.

about 27 miles above the zero level. With this output, it is possible to project speech a distance of 1000 feet under ordinary weather conditions and this has been done at several installations.

On account of the different conditions encountered in installations three types of horns have been used, shown in Fig. 12. Where it is necessary to project the sound to great distances, a tapering wooden

horn is used, of rectangular cross section, $10\frac{1}{2}$ feet long, the walls being stiffened to prevent lateral vibration. For most installations these large horns are not required, and two types of fibre horns are used. One of these is straight in the body, with a flaring open end, while the other used in the control room, is bent.

The grouping of projector units on the volume control switches differs with the type of installation. In outdoor performances, the necessity of correcting the volume in certain directions due to varying winds makes it advisable to group adjacent projectors on a single switch. This is not the case with indoor performances, as no wind effects are possible. Instead, symmetrically placed projectors which will always require equal volume are grouped on a single switch.

Power Supply Equipment

In order to convert the commercial electric power supply to forms suitable for supplying the filament and plate current for the vacuum

Fig. 14a.

tubes in the amplifiers, two types of power equipment have been made available. When the installation is of a size requiring the high-power amplifier, a vacuum tube rectifier taking its supply at 110 or 220

volts, 60 cycles, and delivering 750 volts direct current for the plate circuits is employed. A potentiometer arrangement provides a direct-current supply at 350 volts for the speech-input amplifier tubes. Full wave rectification is obtained and a filter consisting of a large series reactance coil and bridged condensers is used to render the direct-current output suitable for this use. Included in the

Fig. 14b.

power equipment is a step-down transformer for supplying the filaments of the power amplifier. For the larger installations, employing the rectifier, the total power drawn from the commercial supply is 1500 watts.

For installations of a size not requiring the use of the high-power amplifier, a compact motor generator set is provided consisting of a 350-volt d-c. generator driven by a suitable motor, the total power drawn from the supply mains being about 500 watts. A low-voltage generator for supplying direct current at 12 volts for the operation of the amplifier tubes is incorporated in this motor generator set. A filter is necessary and a reactance coil and a 12-volt storage battery is floated across its output. This supplies the transmitters and the tube filaments. The necessary indicating meters are provided on a

meter panel for observing the voltages and currents of all the items of equipment which do not have individual meters associated with them.

Observing System

In addition to the monitoring projector provided in the control room for the guidance of the operator, it has been found necessary in all but the simplest installations to provide observing stations at

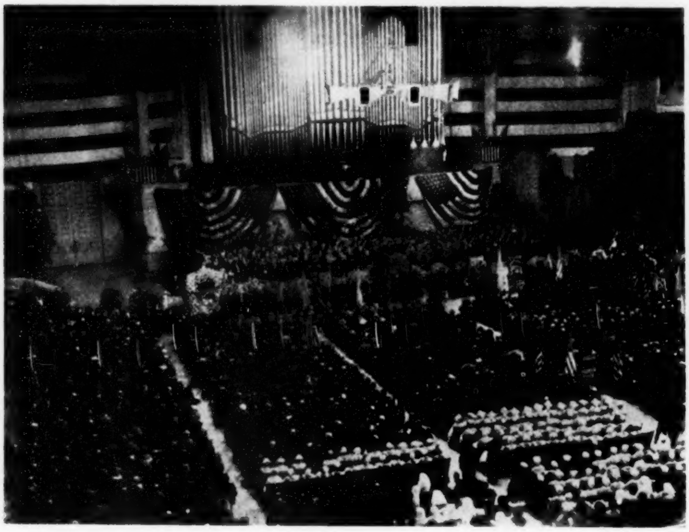


Fig. 15.

various points in the audience. The observers stationed at these points are equipped with portable telephone sets, by which they may immediately communicate with the operator, who is provided with a telephone set consisting of a head receiver and a breast transmitter. The value of these observation stations for regulation of output volume during a program will be apparent.

In the case of an open air performance a variable wind may make it necessary to increase the volume of certain projectors and decrease the volume of others in order to cover the audience uniformly. Without the observers, the control operator would be unable to take care of these changes.

Considerable preparation is required where the equipment is being used for the first time, in order that the performance of the public

address system installation will be of the highest order. Where the acoustic conditions are unfavorable it is necessary to make tests with various arrangements of the projectors, in order to determine the

Fig. 16a.

most satisfactory one. It has been found advisable to carry out the entire program previous to the performance in order that the operating force may become familiar with the sequence.

Conclusion

The usefulness of such systems is very well illustrated by a few of the results which have actually been obtained. Fig. 14 shows a crowd of approximately 125,000 people, every one of whom was able to hear clearly and distinctly all of the words spoken in President Harding's inaugural address in March, 1921. This crowd was relatively small, compared with the crowd which could be accommodated by one of the larger type systems. Some insight into the number of people which could be accommodated can be gained from the fact that such a system will cover comfortably a complete circle whose diameter is 2000 feet when the projector units are placed at the center.

Fig. 16b

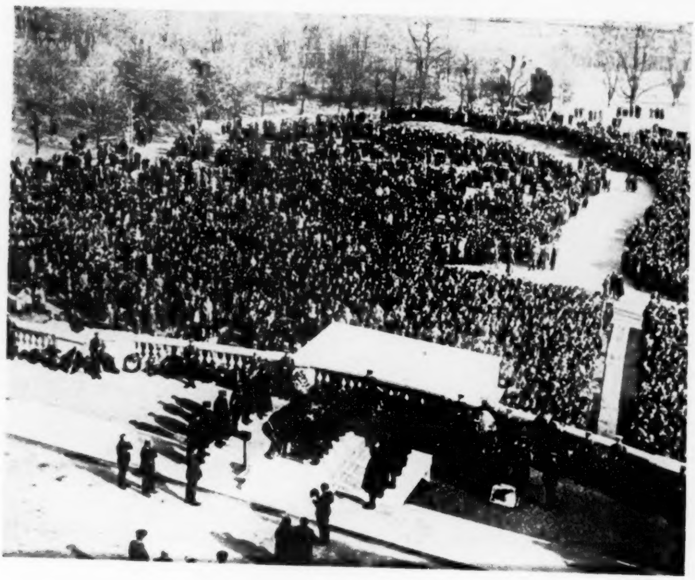


Fig. 17a.

One of the largest and most successful uses to which this equipment has been put took place on Armistice Day in 1921 when 20,000 people in San Francisco, 35,000 people at Madison Square Garden, New York City, and approximately 100,000 people at Arlington Cemetery near Washington joined in the impressive ceremonies which took place at the burial of the Unknown Soldier. Figs. 15, 16 and 17 are views at the three cities, during the ceremonies.

Some of the other uses which have been made of the public address system are the Republican and Democratic Conventions prior to the last presidential election, after-dinner speaking in large ballrooms and in halls where speakers have to address large audiences. There

Fig. 17b.

is one more application of this type of equipment which is gaining rapidly in its use. This last is the application of the speech input equipment to radio broadcasting. The broadcasting of the opera, "Aida," from the Kingsbridge Armory, and of the Philharmonic

Concerts from the Great Hall of the College of the City of New York are two of the successful uses where music and speech were concerned, while broadcasting of the results of the football games from the various distant cities indicates possibilities for the dissemination of interesting information.[9]

The social and economic possibilities of the system are scarcely realized by the public as a whole at the present time, when the method resorted to for reaching large numbers of people is usually the printed word. While this method is effective, it leaves much to be desired in that the personal touch between the man with ideas and the people to receive them is entirely lost. The difficulty for any but those possessing the strongest voices to reach an appreciable number of people at one time has led to a decline in oratory as a means of conveying public messages to large numbers, for it is not always the man with the best ideas or the best ability of presenting them, who is blessed with a powerful voice. A system such as the one which has just been described enables the speaker, even though his voice be relatively weak, to address at one time and in one gathering, several hundred thousand persons, and if the system be used in connection with long distance telephone lines or radio broadcasting, the number which may be reached is increased almost indefinitely. The value of such a situation can hardly be overrated in times of national emergency or stress, when it is necessary for those in responsible positions in the Government to get their message to the people directly.

The development of the apparatus just described has been the result of the efforts of such a large number of investigators working cooperatively that no attempt has been made to acknowledge the individual contributions.

[9] Use of Public Address Systems with Telephone Lines. W. H. Martin and A. B. Clark.

Use of Public Address System with Telephone Lines[1]

By W. H. MARTIN and A. B. CLARK

SYNOPSIS: The combination of the public address system and the telephone lines makes it possible for a speaker to address, simultaneously, audiences located at a number of different places. Such a combination has been used in connection with several public events and a description is given of the system as used on Armistice Day, 1921, when large audiences at Arlington, New York and San Francisco joined in the ceremonies attending the burial of the Unknown Soldier, at the National Cemetery, Arlington, Virginia.

More recently the public address system has been used in conjunction with telephone lines to attain two-way loud-speaker service. This arrangement permits the holding of joint meetings between audiences in two or more locations, separated by perhaps thousands of miles, in such a manner that speakers before each of the audiences can be heard simultaneously by the other audiences. A demonstration of two-way operation was given at the mid-winter convention of the American Institute of Electrical Engineers in February, 1923, and took the form of a joint meeting between 1,000 members in New York and 500 in Chicago.

The electrical characteristics of any telephone line which is to be used in conjunction with loud-speaker equipment must receive special attention. In commercial telephone service the main requirement is understandability, while with the loud-speaker naturalness of reproduced speech is very important. People are accustomed to hearing through the air with very little distortion and naturally expect the same result with loud speakers. A satisfactory line for this purpose must show freedom from transients, echo effects, etc., as well as good uniformity of transmission over the proper frequency range.

The public address system, apparatus and methods has also been applied to radio broadcasting. The combination of the public address system with lines and radio makes it possible for one speaker to address enormous numbers of people located all over the country.—*Editor.*

THE public address system which is described in the preceding paper by I. W. Green and J. P. Maxfield, was developed and first used for the purpose of extending the range of the voice of a speaker addressing an audience. With the aid of this system enormous crowds extending from the speaker's stand to points a thousand feet and more distant have in reality become an audience and have easily understood the speaker whose unaided voice covered only that portion of the crowd within a hundred feet or so from him.

When this system, consisting of a high quality telephone transmitter, distortionless multi-stage vacuum tube amplifiers, powerful loud speaking receivers and projectors, had so shown its capabilities in reproducing speech sounds, a logical extension of its application was to use it with telephone lines. By connecting the transmitting and receiving elements of the public address system through a suitable

[1] Presented at the Midwinter Convention of the A. I. E. E., New York, N. Y., February 14-17, 1923. Published in the *Journal of the A. I. E. E.* for April, 1923.

telephone line a system is provided whereby a speaker can address an audience at a distant point. Also with a complete public address system at the point where the speaker is located, connected by lines to receiving elements of the public address system located at one or more distant points, the speaker is enabled to address a large local audience and to be heard simultaneously by audiences at one or more remote points. This last arrangement was first used on Armistice Day, 1921, when audiences at Arlington, New York and San Francisco joined together in the ceremonies attending the burial of the Unknown Soldier at the National Cemetery at Arlington, Virginia.

By means of the public address system, the meeting of this Institute at New York, at which this paper is presented, is attended and participated in by Institute members at a meeting in Chicago. This is the first occasion on which complete public address systems installed at meetings in two cities have been connected together by telephone lines so that speakers at each meeting address the local and distant audiences simultaneously.

With the transmitting element of the public address system working into the radio transmitter of a broadcasting station and with the receiving elements of the system connected to the output of radio receiving sets, a system is provided whereby a number of people can be reached by each radio receiver.

The combination of these wire and radio communication channels with the elements of the public address system is, therefore, without limit in the number of persons who may be reached simultaneously by one speaker. Such combinations may prove extremely serviceable for occasions of nation-wide interest and importance.

The public address system apparatus has been used not only for transmitting speech sounds but also for music, both vocal and instrumental. The paper[2] describing the public address system has pointed out that the requirements for such a system are that for a wide frequency range it be practically distortionless, that is, transmit and reproduce with equal efficiency all frequencies in that range. This requirement must apply likewise to lines which are used with the loud speaker system. It has been found that a circuit which transmits without material distortion the frequency range from about 400 to 2000 cycles, can be used with the public address system to reproduce speech sounds which are fairly understandable under favorable conditions, although sounding unnatural. In general it is important to extend this range at both ends in order to improve the intelligibility of the sounds and increase the naturalness. For vocal and for some

[2] Green and Maxfield, "Public Address System."

types of instrumental music the melody can be reproduced with the above frequency range, but these tones also are lacking in naturalness. Since some of the musical instruments are used to produce tones three and even four octaves below middle C, it is evident that the proper reproduction of music requires a further extension of the lower limit of the transmitted band than does speech. While the fundamentals of the higher musical tones lie in general in the range mentioned above, it is the harmonics in musical tones which distinguish those produced by different instruments and which give what musicians term "brilliance." The true reproduction of many musical selections requires the distortionless transmission of a frequency band of from about 16 cycles to above 5000 cycles. Many musical selections, however, employ only a part of this range and accordingly can be satisfactorily reproduced by systems not transmitting the whole range. Also, even with slight distortion obtained with somewhat narrower ranges, reproductions may be given which are agreeable to many popular audiences.

Line Requirements

In general the same line requirements which make for satisfactory transmission of speech over commercial telephone circuits also make for satisfactory transmission when telephone circuits are associated with loud speakers. There is this difference however. The loud speakers tend to make the line distortion much more noticeable and serious. Speech transmitted over a particular telephone line is, in general, more difficult to understand when listening to loud speakers than when listening to telephone receivers.

In commercial telephone service the main requirement is intelligibility while, with the loud speaker, the naturalness of the reproduced speech sounds is very important. People are accustomed to hearing transmission through the air with very little distortion and naturally expect the same result with loud speakers.

The above constitute the reasons why, for transmitting voice currents over telephone lines with loud speakers, it is necessary to pay unusual attention to the electrical characteristics of the lines. Evidently when music is to be transmitted, particularly music of a fairly high grade, it is necessary to place even more severe electrical requirements on the lines.

An analysis of what constitutes the electrical requirements of a telephone line which make for good transmission, particularly when loud speakers are employed, will now be given.

In the first place, as explained above, it is essential that a suffi-

ciently broad frequency range be transmitted. As explained in another paper[3] it is not sufficient that a telephone circuit transmit sustained alternating currents within a given frequency range. It must also transmit short pulses of alternating currents within the proper frequency range without introducing oscillations of its own or "transient effects." This requires that loaded circuits for loud speaker use have a high cut-off frequency and hence have the frequencies of the predominant natural oscillations high. It has been found that when the cut-off frequency of loaded circuits is about 5000 cycles, good results are secured with loud speakers.

The two types of telephone circuit which best meet the requirements of transmitting a broad band of frequencies, both when sustained and when applied in short pulses, are non-loaded open-wire lines and extra-light loaded cable circuits. These are suitable for transmission over very long distances. For transmission over short

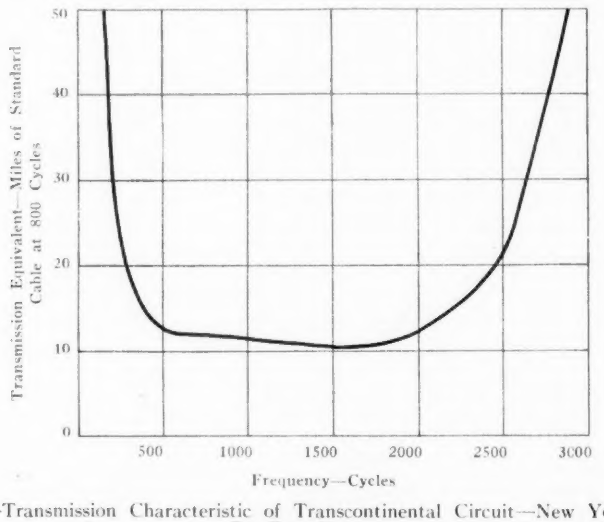


Fig. 1—Transmission Characteristic of Transcontinental Circuit—New York to San Francisco.

distances, say from one point in a city to another point in the same city, non-loaded cable circuits equipped with distortion networks or attenuation equalizers for equalizing the attenuation, give good results.

A good idea of the range of frequencies which can be transmitted

[3] Clark, Telephone Transmission Over Long Cable Circuits, *Journal of A. I. E. E.*, January, 1923. Also Bell System *Technical Journal* for January, 1923.

over high grade telephone circuits can be secured from Fig. 1, which shows the transmission efficiency at different frequencies for the New York-San Francisco circuit. This circuit is a non-loaded No. 8 B. W. G. open wire line equipped with twelve telephone repeaters and is 3400 miles long. Its frequency characteristic meets very well the requirements for easy understanding of voice transmission although it causes some loss of naturalness.

The frequency range which can be transmitted with approximately constant efficiency is limited at the lower end by the fact that composite sets are employed in order to make it possible to superpose direct current telegraph circuits. The elimination of these composite sets would make it possible to improve the transmission of low frequencies and thus improve the operation of the circuit in connection with loud speakers. The resulting improvement, however, would not

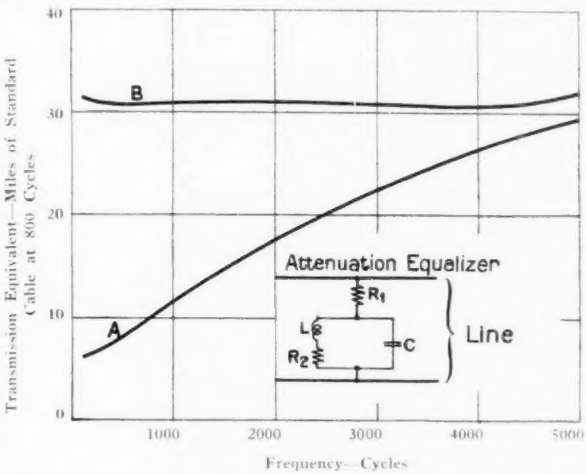


Fig. 2—Transmission Characteristic of No. 19 Gage Non-Loaded Cable Circuit.

A—Without Attenuation Equalizer.
B—With Attenuation Equalizer.

be of importance for commercial telephone service and would render it more difficult to avoid noise on circuits exposed to induction from paralleling power or telegraph circuits.

At high frequencies the range is limited because these same wires are equipped with apparatus to permit super-position of multiplex carrier telegraph circuits above the voice range. This limitation also

is not important for commercial telephone service although it is of importance for loud speaker use. To raise the upper limit of the voice transmission range would require giving up some of these facilities.

Fig. 2 will give an idea of how the distortion introduced by a length of non-loaded cable can be corrected by employing distortion networks or attenuation equalizers. This figure shows the transmission frequency characteristic of about 10 miles of non-loaded No. 19 A. W. G. cable. Curve *A*, in the figure, shows the characteristic when uncorrected, while Curve *B* shows the characteristic for the circuit when equipped with an attenuation equalizer.

After choosing the proper types of telephone circuits for use in connection with loud speakers, there remains to be considered a number of other important matters.

The maintaining of the telephone power within proper limits at different points in the circuit is very important. The power must not be allowed to become too weak, otherwise the extraneous power induced from paralleling circuits would tend to obliterate the telephone transmission. On the other hand, the telephone power must not be amplified to such an extent that the telephone repeaters will be overloaded or severe cross talk be induced into paralleling circuits.

To keep the telephone power throughout the circuit between the above limits, requires careful study and adjustment. For handling regular telephone connections, the circuits are laid out and equipped with repeaters at proper points so that each circuit will be able to handle the varying volumes applied at the terminals when different subscribers are connected without getting into serious difficulties. When loud speakers are employed it is necessary to maintain the volume at the terminals of the toll lines at least within these limits and it is preferable to do somewhat better than this.

With the public address system, the high quality transmitter which picks up the sound at the sending end is usually associated with an amplifier whose adjustment is varied, depending on the output of voice currents from the transmitter. In order to obtain the proper adjustment of this amplifier, it is necessary to have some means for quickly indicating the volume of transmission. For this purpose, there has been developed a device which is called a "volume indicator." This consists of an amplifier detector working into a direct-current meter. With this volume indicator connected across the output of the transmitter amplifier, the volume of transmission delivered to the line is indicated by the deflections on the meter. By adjusting the amplifier, therefore, to keep the deflections of this meter reason-

ably constant at some deflection determined by previous calibration, it is practicable to keep the telephone power within the required limits. Obviously, this same device may also be employed to keep the telephone power constant at any other point in the system.

While the necessity for keeping the power applied to the toll lines within proper limits cannot be over-emphasized, it should also be noted that this is not sufficient. It is also essential that all parts of the toll circuit, including the repeaters, be maintained at prescribed efficiency so that the power levels at all intermediate points in the circuit will also be kept within proper limits. Long telephone lines are designed with special emphasis on this matter of constant efficiency so that, in general, no special precautions are required when using these circuits in connection with loud speakers.

In another paper, [4] the "echo" effects which may occur on long telephone circuits are explained. When setting up two-way circuits for loud speaker use, it is necessary to pay particular attention to effects of this sort. Furthermore, there is another source tending to produce echoes in circuits arranged for two-way use with loud speakers. This is the tendency for the sound delivered from the loud speaker projectors to enter the sensitive transmitter and be returned to the distant end of the circuit as an echo. Owing to the relatively slow velocity of transmission of sound through air the lag in such an echo may be great enough to be serious, although the line is a short one with high transmission velocity. It is, therefore, evident that this coupling through the air between the loud speaker projector and the transmitter must be kept small. If a very sensitive transmitter arranged so that a speaker may stand several feet away from it is employed, this problem becomes even more difficult.

There is one thing more that remains to be considered: the necessity for special operation. When a large number of people are assembled at some point to hear an address delivered at a distant point, it is evident that delay in establishing the connections cannot be tolerated. It is, therefore, necessary to establish such connections ahead of time and it is usually also necessary to set up spare circuits for use in case of failure of the regular circuits. A special operating force is required for checking up the circuits, establishing the connections when required, and making the necessary adjustments. Rehearsals are necessary on important occasions to insure proper functioning of the circuits and proper co-ordination of the handling of the circuits with the programs at different points.

[4] Clark, *loc. cit.*

Typical Circuit Combinations of Public Address System and Lines

Following are a number of typical combinations of the public address system and telephone lines. The combinations by means of which one-way service may be rendered, are given first, following which certain combinations for giving two-way service are discussed.

By one-way service is meant service in which no provision is made for anyone in the distant audience to talk to the place where the speaker is located. Two-way service provides for speakers at either of two or more points addressing all of the other points. This is similar to the two-way service rendered by regular telephone circuits.

Fig. 3 shows the circuit arrangement which would be used when a speaker at one point in a city, for example, at his office, is to address

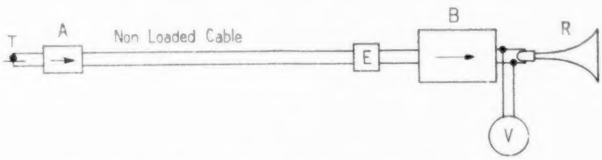


Fig. 3—One-Way Connection to Point in Same City.

an audience at another point in the same city. A high quality close talking transmitter *T*, together with a fixed gain single-stage amplifier *A*, are provided at the point where the speaker is located. This combination is designed to deliver to the line the same amount of power as a commercial type substation set. Connecting this point with the point at which the audience is gathered is a non-loaded cable circuit. To correct for the distortion in this cable circuit, an attenuation equalizer *E* is provided. The apparatus at the point where the audience is located is the equipment of the public address system without the transmitter and its associated amplifier. In Fig. 3, *B* is the amplifier for delivering sufficient power to the group of loud speaker projectors indicated by *R*. A volume indicator *V* associated with the amplifier *B* is used in maintaining constant the volume of sound delivered from the projectors.

Fig. 4 shows the circuit combination required when a connection is to be established to a distant city where the loud speaking receivers are located. In the city where the speaker is located, connection is made to the toll office by means of a non-loaded cable circuit equipped with an equalizer similar to Fig. 3. A volume indicator V_1 is associated with the amplifier C_1 at the toll office to enable proper adjust-

ment of amplifier C_1 to be made so that the power delivered to the toll line will be within the proper limits. As explained above if the volume at the toll office is allowed to become too great, the telephone repeaters on the toll line will be over-loaded and serious distortion will result, while if the volume is allowed to become too weak, extraneous noise

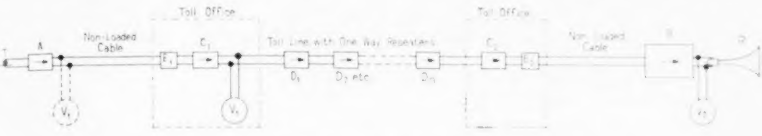


Fig. 4—One-Way Connection to Point in Distant City.

and crosstalk will tend to obliterate the direct transmission. If a distant talking transmitter is used for the speaker, a multi-stage adjustable amplifier is associated with it. In this case the volume indicator is located at the output of this amplifier as shown by the dotted lines in Fig. 4. When the volume indicator is employed at this point it is necessary to take into account the loss introduced by the non-loaded cable and the equalizer E_1, together with the gain of the repeater C_1, in order to deliver volume within proper limits to the toll line. The toll line, shown equipped with repeaters D_1, D_2, etc.,

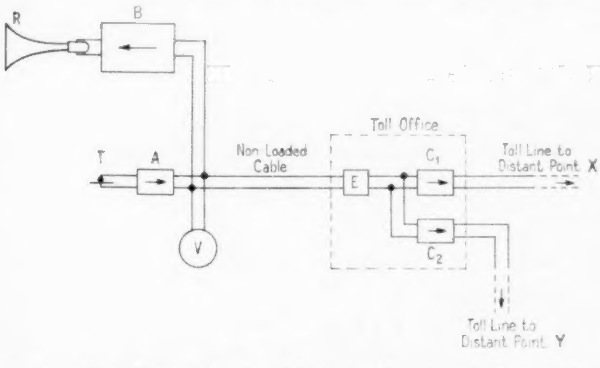


Fig. 5—One-Way Connection for Addressing Local and Distant Audiences Simultaneously.

extends to the toll office in the distant city. At this point the amplifier C_2 is located, together with another equalizer E_2, for correcting the distortion in the local non-loaded cable circuit. The apparatus at the point where the audience is located is similar to that shown in Fig. 3.

Fig. 5 shows the circuit combination employed when a local address is to be given, while at the same time the same address is delivered to one or more distant points. In order to allow the local audience to hear the address by means of the loud speakers, the power amplifier *B* supplying energy to these is bridged across the output of the amplifier *A* associated with the transmitter *T*. A volume indicator *V*, connected across the circuit at the point where the bridge is made, makes it possible to maintain constant volume both for the local loud speakers and for the transmission applied to the toll lines by suitable adjustment of amplifier *A*. At the toll office means are indicated for connection to two distant points *X* and *Y*. Owing to the fact that amplifiers C_1 and C_2 are one-way devices, no inter-actions can occur between lines *X* and *Y* or between these lines and the local loud speaking system. The arrangements for reaching the distant points *X* and *Y* are similar to the one illustrated in Fig. 4.

All of the circuit arrangements which have so far been described are arranged simply so that a speaker may address one or more local or distant points. When it is desired that the speaker and the audience at the sending end also be able to hear a speaker at the distant point, more complicated arrangements are required.

Fig. 6 shows a circuit arranged for such two-way service, the line being operated on the four-wire principle, *i. e.*, two separate transmis-

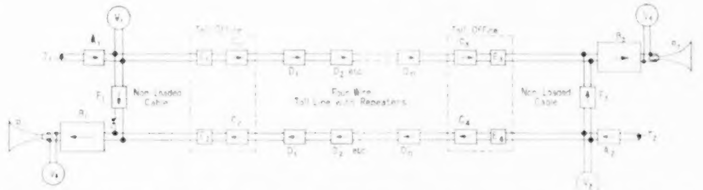


Fig. 6—Two-Way Four-Wire Connection for Addressing Local and Distant Audiences

sion paths are provided, one for transmission in each direction. The circuits connecting transmitter T_1 with the projector group R_2 and transmitter T_2 with the projector group R_1 are similar to the circuit in Fig. 4. By-pass connections F_1 and F_2 are added at the two ends which allow part of the output of each transmitter to pass into the local loud speakers. These by-pass connections are so arranged that transmission can pass only in the proper direction. Two volume indicators are provided at each end. Referring to the left-hand terminal, volume indicator V_1 is provided to insure that power is supplied to the toll line within the proper limits of volume, as explained above.

V_3 is provided to facilitate adjustment of the by-pass circuit F_1 and of amplifier B_1 so as to deliver proper volume from R_1 both for the local talking and for the reception of the addresses from the distant end of the circuit. The volume indicators V_2 and V_4 at the right-

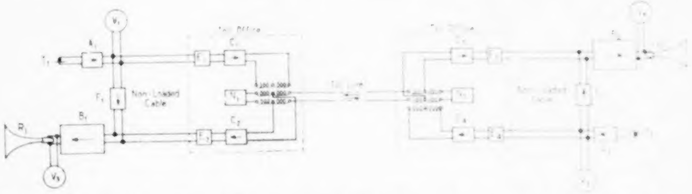


Fig. 7—Two-Way Two-Wire Connection for Addressing Local and Distant Audiences.

hand end of the circuit have functions similar to those of V_1 and V_3 respectively.

Fig. 7 is similar to Fig. 6 with the exception that the toll line is of the two-wire type. At each end of the toll line, which may, or may

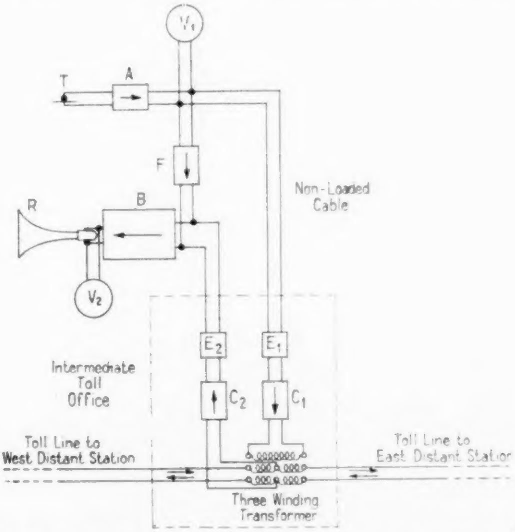


Fig. 8—Arrangement for Connecting Third Point to Circuit of Fig. 7.

not, contain two-way repeaters, transformers and networks N_1 and N_2 are placed for converting the two-wire circuit into a four-wire circuit. The equalized cable circuits at the two ends thus form two

sides of short four-wire circuits. The conditions of balance between the networks and the toll lines prevent more than a very small amount of the direct transmission from each local transmitter from entering the local loud speading receiver circuit at the points where the local circuits connect to the toll line. Practically all of the transmission from transmitter T_1 to projector group R_1 and from transmitter T_2 to projector group R_2 is delivered through the adjustable by-pass circuits F_1 and F_2, respectively.

For connections requiring to and fro conversations between three or more points, all of which may be equipped with loud speakers, intermediate points may be connected to a two-wire telephone circuit by employing the arrangement shown in Fig. 8. A three-winding transformer is inserted in the toll line which is so constructed that the impedance which it introduces into the circuit is small enough to avoid a serious irregularity. Talking currents are put out on the toll line through this transformer. The received transmission is obtained from a high impedance bridge across the midpoints of two of the windings of the three-winding transformer. Amplifiers C_1 and C_2 introduce sufficient gain to overcome the losses due to the inefficient coupling with the telephone line. The rest of the circuit at the intermediate point is the same as Figs. 6 and 7, the local speaker being heard by his own audience by means of transmission delivered through by-pass F. A modification of the arrangement of Fig. 8 can, of course, be used with a four-wire toll circuit.

Arrangements for Armistice Day, 1921

Fig. 9 shows the circuit which was employed on Armistice Day, 1921, when audiences of 100,000 people at Arlington, 35,000 people at New York and 20,000 people at San Francisco, joined in the services at the burial of the Unknown Soldier. This was the first time that audiences at more than one distant point were simultaneously addressed from one point by means of the public address system.

At Arlington three different transmitters T_2, T_3 and T_4 were used for the different parts of the ceremonies. T_2 was used for the musical selections, T_3 for the speeches made in the amphitheatre, and T_4 for the speeches at the grave of the Unknown Soldier. Another transmitter T_1 was provided for the use of an announcer who kept the audiences at New York and San Francisco advised of the proceedings. The speech currents leaving these transmitters were brought up to moderate volume by means of amplifiers A_2 and A_1, the former taking care in turn of the three different transmitters employed during the ceremonies.

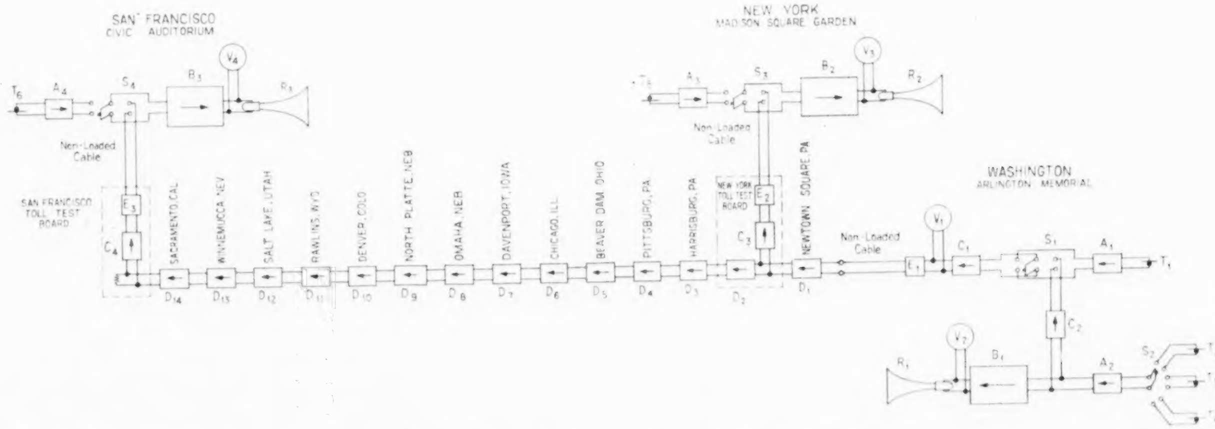


Fig. 9—Circuit Used for Ceremonies on Armistice Day, 1921.

The voice currents from the transmitters which were employed for the ceremonies, after passing through amplifier A_2, separated into two branches, one branch going to the local amplifier B_1, which supplied the local loud speakers R_1, the other going to the telephone circuit through amplifier C_2, switch S_1 and amplifier C_1. The switch S_1 was provided for connecting either the announcing transmitter T_1 or one of the transmitters for picking up the ceremonies to the end of the toll line. V_1 and V_2 are volume indicators, V_1 being employed to indicate that the proper power was being put into the toll line, while V_2 furnished an indication of the volume which was being delivered by the projector group R_1. During the ceremonies the amplifier C_1 was continuously adjusted so as to deliver proper volume to the long distance telephone circuit, the volume indicator V_1 making it possible to keep the volume applied to the toll line within close limits. At the same time independent adjustments were made of the amplifier B_1 to take care of the varying conditions introduced by the different talking conditions as well as the varying conditions introduced by shifting of the crowds listening to the ceremonies.

After leaving the amplifier C_1 at Arlington, the voice currents first passed through a non-loaded section of cable whose distortion was corrected by equalizer E_1. A non-loaded 8-gauge open-wire circuit carried the voice currents to New York City. At this point, the circuit again branched, one branch delivering a part of the voice currents to the apparatus at Madison Square Garden, the other branch going to San Francisco over one of the non-loaded No. 8-gauge transcontinental circuits. The arrangements employed at Madison Square Garden and at the Civic Auditorium in San Francisco were similar, switches being provided at each point to connect to the projector groups the circuit from Arlington or from the local transmitter.

The difficulties involved in transmitting voice currents for the first time to loud speaker installations at distant points, as well as the great importance of the occasion, made it necessary to take elaborate precautions in order to insure the success of the undertaking. The long distance telephone circuits were carefully inspected ahead of time and all of the amplifiers and other apparatus employed were subjected to numerous careful tests. For checking the complete circuit, alternating currents of different frequency were applied at Arlington and measured simultaneously at New York and San Francisco. The curve on Fig. 1 was obtained from the results of one of the measurements made on this occasion.

To guard against possibility of failure of the circuits, emergency circuits were provided, these emergency circuits taking different

routes wherever possible. Fig. 10 shows the network of long distance circuits which was set up for this occasion. The solid lines in this figure indicate telephone circuits while the broken lines indicate telegraph circuits. The latter were for the purpose of transmitting orders

Fig. 10—Telephone and Telegraph Lines Used on Armistice Day, 1921

between different units of the operating organization at different points.

At Arlington the nature of the ceremonies and the place in which they were held presented many difficulties from the acoustic standpoint. The main addresses were made in an open amphitheatre surrounded by a double colonnade of marble. The platform on which the speakers were located was partially covered by a marble arch. The floor of the amphitheatre is of cement on which are arranged marble benches. Temporary seats also were placed on top of the colonnade. During the ceremonies large crowds surrounded the amphitheatre on all sides. The arrangement of the amphitheatre and the surroundings is shown by Fig. 11.

In order that the crowds outside of the amphitheatre might hear the speakers, loud speaking receivers and their associated projectors were placed on top of the colonnade. They were arranged in four groups as shown on Fig. 11, the projectors referred to, being numbered from 1 to 21 inclusive. Those in the east group were on top of the structure forming the main entrance to the amphitheatre. The

projectors were carefully directed to cover uniformly the area around the amphitheatre and were supplied with sufficient power so that the speaker could be heard for at least a thousand feet from the outside of the amphitheatre. It was found, however, that while these projectors are highly directive, some of the sound from them could be heard inside the amphitheatre. This sound leakage at the western side was particularly serious because of the fact that it reached the rear seats inside of the amphitheatre sufficiently far enough ahead of the corresponding sounds directly from the speaker to be noticeable.

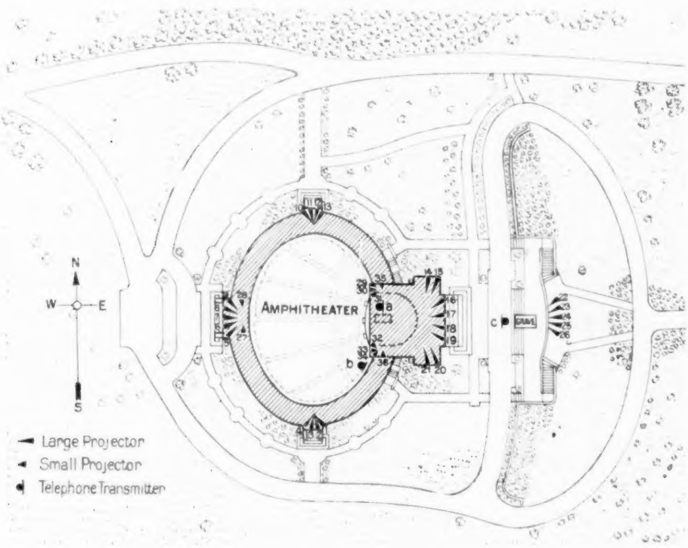


Fig. 11—Arrangement of Projectors at Arlington Amphitheater

To overcome this, the small projectors 29, 31, 32 and 34, placed on top of the arch over the platform, were directed at the rear seats and given sufficient volume output to overcome the sound reaching these seats from the loud speakers on the colonnade.

The adjustment of the power to these small projectors required great care because if given too great volume, bad reflections would be set up in the amphitheatre. On the other hand if this volume were not great enough, the outside projectors would cause serious interference. The small projectors 27 and 28 were used to overcome the sound leakage effects on the top of the west side of the colonnade.

The projectors 35 and 36 covered the top of the colonnade on the east side.

Fig. 11 shows also the location of the three transmitters used during the ceremonies, *a*, on the platform for the speakers, *b*, in front of one of the boxes in which were placed the singers and behind which was located the band, and *c*, at the grave. When the transmitter at the grave was tested it was found that serious interference was obtained between the speaker's voice and the sound from the projectors 16 to 19 inclusive. For the ceremonies at the grave, therefore, these loud speakers were disconnected and those numbered 22 to 26 used instead. Also in order to properly cover the inside of the amphitheatre during the ceremonies at the grave, the small projectors 30 and 33 were used. These were located on the arch over the platform and were directed at the front seats in the amphitheatre.

The projectors were divided up into a number of small groups and so connected that the volume of sound delivered by each group could be varied without affecting the other groups. This was necessary in arriving at the power to be delivered by each projector to give uniform distribution and to avoid interference between different groups.

By means of these arrangements all parts of the ceremonies were carried to all parts of the audience at the National Cemetery and were also delivered by means of the lines to the audiences in the distant cities.

At New York, a group of fifteen loud speakers was used in Madison Square Garden to satisfactorily reach all parts of the audience and a group of twenty-one loud speakers was suspended outside the building for the outside audience. At San Francisco, ten loud speakers were used in the Civic Auditorium and seven outside.

Use of Public Address System Apparatus with Radio

When radio broadcasting came into general use, the apparatus and methods which had been developed for the public address system were applied to this new field as it also demands high quality reproduction for speech and music. The transmitters and amplifiers associated with them in the public address system are used in radio broadcasting studios for delivering speech frequency electrical power to the radio transmitter. Loud speaking receivers and amplifiers for delivering sufficient power to operate them are used with many of the radio receiving sets.

The methods which have been employed to connect public address system transmitters with toll lines are being used for the broadcasting by radio of speeches and music given at points remote from the

radio station. In such cases the transmitter and its associated amplifier are operated and controlled in the same way as described above for toll lines. In some cases the radio station is in the same city as the place where the speech or music is given and in other cases the two have been in different cities. In the first case the output of the transmitter amplifier is carried to the radio station over non-loaded cable circuits which are equalized by means of distortion correction networks to have uniform efficiency over a wide frequency range, in some cases up to 5000 cycles. Where the two points are in different cities, the non-loaded cable circuit goes to the toll office and there is connected to the toll lines which are operated in the same manner as described above for loud speaker use.

For some of the higher grade music, such as that given by symphony orchestras, the less efficient, but slightly higher quality condenser type transmitter has been used instead of the double button carbon type. This requires the use of an additional two stage amplifier in front of the regular three stage transmitter amplifier.

The output of the transmitter amplifiers is controlled with the aid of a volume indicator bridged across the output terminals of the amplifier. For best results, particularly in reproducing music, it is necessary to adjust the gain of these amplifiers to compensate partially for the large range in the volume of the music. If the amplifiers are set high enough in gain to send through the low passages of the music with sufficient volume so that it will override the static and the interference from other sending stations, the loud parts of the music will seriously overload the radio transmitter system, unless it is of very large capacity, and will in general overload the receiving sets. Furthermore putting out these loud parts at the same relative level with respect to the low passages as they are given by the orchestra, makes the interference between radio stations more serious. In some orchestral concerts the power amplification of the transmitter amplifier has been adjusted over a range of more than a hundred to one, these changes being made, however, so that they were not noticed by those listening to the concert by radio.

Proper volume control is very important in picking up such music for radio broadcasting. The lack of such control is responsible for many of the poor results that are being obtained. In this connection, the location of the transmitter with respect to the various instruments in the orchestra or smaller combination of instruments, so as to maintain in the reproduced music the proper balance between the several parts is, of course, of great importance.

An interesting illustration of the combination of the public address

system, telephone lines and radio broadcasting was used in connection with reporting a football game played in Chicago in the fall of 1922. By means of high quality transmitters and amplifiers located at the football field, announcements of the plays and the applause of the spectators were delivered to a circuit extending to the toll office in Chicago. This circuit was connected there to a toll line to New York where it delivered the telephonic currents to a radio broadcasting transmitter. In Park Row, in New York City, was located a truck on which was mounted a radio receiving set arranged to operate a public address system. By this means the reports of the plays of the football game in Chicago were delivered to a large crowd in the streets of New York.

The Contributors to this Issue

Ray S. Hoyt, B.S. in electrical engineering, University of Wisconsin, 1905; Massachusetts Institute of Technology, 1906; M.S., Princeton, 1910. American Telephone and Telegraph Company, Engineering Department, 1906–07. Western Electric Company, Engineering Department, 1907–11; American Telephone and Telegraph Company, Engineering Department, 1911–19; Department of Development and Research, 1919—. Mr. Hoyt has made contributions to the theory of transmission lines and associated apparatus, and more recently to the theory of crosstalk and other interference.

Arthur F. Rose, B.S., Colorado College, 1914. Engineering Department, American Telephone and Telegraph Company, 1914–19; Department of Operation and Engineering, 1919—. Mr. Rose's work has been largely in connection with carrier, general transmission and radio matters.

Edward B. Craft, Engineering Department, Western Electric Company, Chicago, 1902–7; Development Engineer, Western Electric Company at New York, 1907–18; Assistant Chief Engineer, 1918–22; Chief Engineer, 1922—.

Mr. Craft has been in charge for the Western Electric Company of the entire program of development work along machine switching lines. He has written papers on radio telephony and electrical communications in war time.

L. F. Morehouse, B.S., University of Michigan, 1897; A.M., 1904; instructor in physics, University of Michigan, 1902–4; junior professor of Electrical Engineering, 1904–6; Transmission Engineer for Western Electric Company in Europe, 1906–9; Equipment Engineer, American Telephone and Telegraph Company, 1909–19; Equipment Development Engineer, 1919—. Mr. Morehouse is responsible for the development and research work affecting the central office and substation equipment of the Bell System, and for the establishment of such standards of central office and substation equipment as are warranted by the results of this development and research work.

H. P. Charlesworth, B.S., Massachusetts Institute of Technology, 1905; Engineering Department, American Telephone and Telegraph Company, 1905–19; Equipment and Transmission Engineer, Department of Operation and Engineering, 1919; Plant Engineer, 1920—. Mr. Charlesworth has had broad experience in the development of telephone equipment and with traffic conditions and the standardization of operating methods and practices.

R. V. L. Hartley, A.B., Utah, 1909; B.A., Oxford, 1912; B.Sc., 1913; instructor in physics, Nevada, 1909–10; Engineering Department, Western Electric Company, 1913—. For some time Mr. Hartley has been closely connected with the development of carrier current, telephone repeater, and telegraph systems.

I. W. Green, Graduate in Applied Electricity, Pratt Institute, 1905; Engineering Department, New York and New Jersey Telephone Company, 1905–7; Engineering Department, New York Telephone Company, 1907–19; Department of Development and Research, American Telephone and Telegraph Company, 1920—. Mr. Green has been engaged on foreign wire relations and more recently on the design of subscribers' apparatus and special problems, such as the public address system.

W. H. Martin, A.B., Johns Hopkins University, 1909; S.B., Massachusetts Institute of Technology, 1911. American Telephone and Telegraph Company, Engineering Department, 1911–19; Department of Development and Research 1919—. Mr. Martin's work has related particularly to loading, quality, and transmission of telephone sets and local circuits.

Joseph P. Maxfield, S.B. in Electrochemistry, Massachusetts Institute of Technology, 1910, assistant instructor in Electrochemistry, 1910–12; instructor in Physics, 1912–14; Engineering Department, Western Electric Company, 1914—. In addition to being closely in touch with the development of loud speaker systems, Mr. Maxfield's work has covered the fundamental study of microphonic contacts.

A. B. Clark, B.E.E., University of Michigan, 1911; American Telephone and Telegraph Company, Engineering Department, 1911–19; Department of Development and Research, 1919—. Mr. Clark's work has been connected with toll telephone and telegraph systems.

The Bell System Technical Journal

Devoted to the Scientific Engineering Aspects of Electrical Communication

July, 1923

Transient Oscillations in Electric Wave-Filters

By JOHN R. CARSON and OTTO J. ZOBEL

I. Introduction

THE electric wave-filter has been very fully discussed with respect to its remarkable steady-state properties.[1] In the present paper it is proposed to give the results of a fairly extensive theoretical study of its behavior in the transient state. This study is of particular interest and importance in connection with the wave-filter, because, as we shall find, its remarkable selective characteristics are peculiarly properties of the steady state and become sharply defined only as the steady state is approached. To this fact, it may be remarked in passing, is to be ascribed the uniform failure of wave-filters to suppress irregular and transient interference, such as "static," in anything like the degree with which they discriminate against steady-state currents outside the transmission range. This limitation is common to all types of selective networks and restricts the amount of protection it is possible to secure from transient or irregular interference. In fact the general conclusions of the present study are applicable to all types of selective circuits.

In the present paper the discussion will be principally concerned with the following phases of the general problem.

1. *The indicial admittances of a representative set of wave-filters.* The *indicial admittance*, as explained below, is equal to the current, expressed as a time function, in response to a uniform steady e.m.f. of unit value, applied to the network at time $t=0$. It has been shown in previous papers that a knowledge of the indicial admittance of an invariable network completely determines its behavior, both in the transient and steady state, to all types of applied forces. Its determination is therefore fundamental to the whole problem.

2. *The mode in which the steady-state is built up after a sinusoidal voltage within the frequency transmission range is applied to the wave-*

[1] Physical Theory of the Electric Wave-Filter, G. A. Campbell, *B. S. T. J.*, Nov., 1922; Theory and Design of Uniform and Composite Electric Wave-Filters, O. J. Zobel, *B. S. T. J.*, Jan., 1923.

filter. Formulas are deduced and a set of representative curves computed and plotted which show the dependence of the building-up process on the constants and number of sections of the filter and the frequency of the applied e.m.f. The outstanding deduction from this phase of the problem is that as the selectivity of the filter is increased either by narrowing the transmission band or increasing the number of sections, the time required for sinusoidal currents to build up is proportionally increased. This fact, it may be remarked, sets a theoretical limit to the amount of selectivity which can be employed in communication circuits.

3. *The character and duration of the transient current when a sinusoidal voltage outside the frequency transmission range is applied to the filter.* It will be found that in this case a transient disturbance penetrates the filter which is enormous compared to the final steady state. The magnitude of this disturbance decreases very slowly with the number of filter sections and its duration increases therewith. This phenomenon is an important special case of the general limitations of the selectivity of the filter in the transient state.

4. *The energy which penetrates through selective circuits from random interference.* The energy spectrum of random interference, that is, interference from random disturbances is discussed and a formula is deduced which defines the *figure of merit* of a selective circuit with respect to random interference. This formula leads to general deductions of practical importance regarding the relative merits of selective networks in the transient state and their inherent limitations. It also provides a method for experimentally determining the spectrum of random interference.

Unfortunately the complexity of transient phenomena is such as to absolutely require a large amount of mathematical analysis. Consequently, while the mathematics has been relegated as far as possible to Appendices, a considerable amount necessarily appears in the text. The writers, however, have endeavored to emphasize the physical significance of the mathematics and have included only that which is absolutely essential to an understanding of the problem and the appropriate methods of attack.

In order to keep the analysis within manageable limits and in a form to admit of relatively simple and instructive interpretation, the formulas will be restricted for the most part to non-dissipative filters and the effects of terminal reflections will be ignored.[2] These

[2] The general solution for the case of arbitrary terminal impedances is given in Appendix IV and briefly discussed.

restrictions are desirable on their own account, because the selective properties, both in the transient and steady-state, are isolated and exhibited in the clearest manner when the disturbing effects of dissipation and reflections are absent. As regards dissipation, its effect is usually small for filters of ordinary length and, as regards transient phenomena, is always of such a character as to require no essential modification of the conclusions reached from a study of the ideal non-dissipative filter. In fact the conclusions reached in this paper regarding the inherent limitations of selective circuits in the transient state are conservative.

II. General Theory and Formulas

Before taking up the investigation of wave-filters it is necessary to write down the fundamental formulas of electric circuit theory, which are required in the analysis, and briefly discuss their application to the investigation of transient phenomena in networks in general. The theory and calculation of electrical networks may be approached in a number of ways, as for example, from the Fourier integral.[3] Perhaps the simplest way, however, is to base the theory on the fundamental formulas

$$I(t)=\frac{d}{dt}\int_0^t f(t-y)A(y)dy, \qquad \text{I}$$

and

$$1/pZ(p)=\int_0^\infty e^{-pt}A(t)dt. \qquad \text{II}$$

In these formulas $I(t)$ is the current (expressed as an explicit time function) in any branch or mesh of an electric network which flows in response to the electromotive force $f(t)$ which is applied to the network at time $t \geqq 0$ in the same or any other branch or mesh of the network. The function $A(t)$ is a characteristic function of the constants and connections of the network only which may be termed the *indicial admittance* or the *Heaviside Function*. Its physical significance may be inferred by setting $f(t)=1$, whence it follows that $I(t)=A(t)$. That is to say *$A(t)$ is equal to the current in response to a "unit e.m.f." (zero before, unity after time $t=0$).*

In the following we shall be principally concerned with the case when the applied electromotive force is sinusoidal. To deal with this case we set $f(t)=\sin(\omega t+\Theta)$ and equation I becomes

$$I(t)=a(\omega,t)\sin(\omega t+\Theta)+b(\omega,t)\cos(\omega t+\Theta) \qquad \text{III}$$

[3] The Solution of Circuit Problems, T. C. Fry, *Phys. Rev.*, Aug., 1919.

where, denoting $d/dt\, A(t)$ by $A'(t)$,

$$a(\omega, t) = A(o) + \int_0^t \cos \omega y\, A'(y)\, dy$$

and IV

$$b(\omega, t) = -\int_0^t \sin \omega y\, A'(y) dy.$$

The ultimate steady-state amplitudes are evidently the limits of the foregoing as t approaches infinity. Thus if we write the steady-state current as

$$I = \alpha(\omega) \sin (\omega t + \Theta) + \beta(\omega) \cos (\omega t + \Theta),$$

then

$$\alpha(\omega) = A(o) + \int_0^\infty \cos \omega y\, A'(y) dy$$

and V

$$\beta(\omega) = -\int_0^\infty \sin \omega y\, A'(y) dy$$

For the derivation and a fuller discussion of the foregoing formulas the reader is referred to "Theory of the Transient Oscillations of Electrical Networks and Transmission Systems," Proc. A. I. E. E., March, 1919.

In the majority of the more important selective networks $A(o) = 0$; that is to say the initial value of the current is zero and the current in response to the applied sinusoidal voltage of the frequency $\omega/2\pi$ is built up entirely from the progressive integrals

$$a(\omega, t) = \int_0 \cos \omega y\, A'(y) dy$$

and

$$b(\omega, t) = -\int_0^t \sin \omega y\, A'(y) dy$$

in accordance with formula IV. The derivative $A'(t) = d/dt\, A(t)$ of the indicial admittance which appears in the integrals will be termed the *impulse function* of the network to indicate its direct physical significance; it is equal to the current in response to a "pulse" of infinitesimal duration and moment (or time integral) unity, or, stated in the terminology of the radio engineer, it is equal to the response of the network to "shock-excitation." These formulas therefore establish a definite quantitative relation between the selective properties of the network and its response to "shock-excitation"; a relation which

is of great importance in understanding and interpreting the behavior of selective networks to transient disturbances.[4]

The indicial admittance $A(t)$ is calculable from and may be regarded as defined by the very compact formula II.[5] In this equation $Z(p)$ is the operational impedance of the network. It is derived from the differential equations of the problem by replacing the differential operator d/dt by the symbol p, thus formally reducing the equations to an algebraic form from which the ratio $1/Z(p)$ of the current to electromotive force is gotten by ordinary algebraic processes. $Z(p)$ will involve the constants and connections of the network and will depend, of course, on the mesh or branch in which the electromotive force is inserted and that in which the required current is measured.

The procedure in formulating the transient behavior of networks is as follows. Derive the operational impedance $Z(p)$ as stated above. With $Z(p)$ formulated, the corresponding indicial admittance $A(t)$ is determined by the integral equation II. The appropriate methods of solution of the integral equation are briefly discussed in "The Heaviside Operational Calculus." Sometimes the solution can be recognized by inspection as in the case of the low pass wave-filter. Otherwise the procedure in general is to expand $1/Z(p)$ in such a form that the individual terms of the expansion are recognizable as identical with infinite integrals of the required type. Two expansions of this kind lead to the Heaviside Expansion and power series solution, respectively. The appropriate form of expansion depends on the particular problem in hand and often calls for considerable ingenuity and experience. An excellent illustration of the appropriate process is furnished by the detailed derivation [6] of the indicial admittance of the band pass filter which is rather intricate.

In connection with the problem of the energy absorbed from forces of finite duration, and from random interference, the following formulas are required, of which VIII and IX are original and hitherto unpublished. Formula X, which is a special case of VIII and IX was derived by Rayleigh (*Phil. Mag.*, Vol. 27, 1889, p. 466), in connection with an investigation of the spectrum of complete radiation.

If an applied force $f(t)$ exists only in the finite time interval $0 \leqq t \leqq T$, during which it has a finite number of discontinuities and a

[4] It may be noted in passing that these formulas show the futility of attempting, as so many inventors have done in connection with the problem of protection from "static" disturbances, to design a circuit, which, in the language of patent specifications, shall be unresponsive to shock excitation while at the same time shall be sharply responsive to sustained forces.

[5] The Heaviside Operational Calculus, J. R. Carson, *B. S. T. J.*, Nov., 1922.

[6] See Appendix I.

finite number of maxima and minima, it is representable by the Fourier integral

$$f(t) = \frac{1}{\pi}\int_0^\infty |F(\omega)| \cos [\omega t + \Theta(\omega)]\, d\omega, \qquad \text{VI}$$

where

$$|F(\omega)|^2 = \left[\int_0^T f(t) \cos \omega t\, dt\right]^2 + \left[\int_0^T f(t) \sin \omega t\, dt\right]^2. \qquad \text{VII}$$

Let this force be applied to a network in branch 1 and let the resultant current $I_n(t)$ be measured in branch n. Let the steady-state transfer impedance at frequency $\omega/2\pi$ be denoted by $Z_{1n}(i\omega)$ and let $z_n\,(i\omega)$ and $\cos \Theta_n$ denote the impedance and power factor of branch n at frequency $\omega/2\pi$. It may then be shown that

$$W' = \int_0^\infty [I_n(t)]^2 dt = \frac{1}{\pi}\int_0^\infty \frac{|F(\omega)|^2}{|Z_{1n}(i\omega)|^2} d\omega \qquad \text{VIII}$$

and, as special cases,

$$\int_0^\infty [A'_{1n}(t)]^2 dt = \frac{1}{\pi}\int_0^\infty \frac{d\omega}{|Z_{1n}\,(i\omega)|^2}, \qquad \text{IX}$$

and

$$\int_0^T [f(t)]^2 dt = \frac{1}{\pi}\int_0^\infty |F(\omega)|^2 d\omega. \qquad \text{X}$$

The total energy W, absorbed by branch n from the applied force is given by

$$W = \frac{1}{\pi}\int_0^\infty \frac{|F(\omega)|^2}{|Z_{1n}\,(i\omega)|^2} |z_n\,(i\omega)| \cos \Theta_n \cdot d\omega. \qquad \text{VIIIa}$$

Comparison of the formulas for W′ and W shows that, if the branch n is a simple series combination of impedance elements, W′ *is the energy absorbed by a unit resistance element in branch n from the applied force* $f(t)$.

In the subsequent discussion of the behavior of selective circuits to random interference and applied forces of finite duration, W′ of formula VIII will be taken, therefore, as a measure of the energy absorbed by the receiving branch or element. Similarly formula IX measures the energy absorbed when the applied force is impulsive. The application of formula VIII rather than VIIIa, when they differ except for a constant, is justified because we are concerned with the energy absorbed by a receiving element proper, which can be represented by a simple resistance.

The advantage of formula VIII, in addition to its simplicity, resides in the fact that the right hand side is usually quite easily computed,

since the integrand is everywhere positive, and this without any explicit reference to the transient phenomena themselves. Formula IX is of particular importance, because, as will be shown in a subsequent part of this paper, it represents, except in limiting cases, the relative amount of energy absorbed from random interference.

III. The Indicial Admittances of Wave-Filters

We are now in possession of the necessary formulas and mathematical processes for investigating the behavior of wave-filters in the transient state. We shall first write down the indicial admittances of the representative types investigated, their derivation being discussed in Appendix I. The formulas given for the low pass and the high pass are exact, while those of the band pass filters are approximations based on the assumption that the transmission band-width is small compared with the "mid-frequency" of the transmission band. They are therefore formally restricted in their application to "narrow band" filters. The analysis of the exact formula, given in Appendix I, shows, however, that the deductions drawn from the approximate formulas of the text are quite generally applicable without errors of any practical consequence to band pass wave-filters, even when the transmission band is relatively wide. These questions are fully discussed in the Appendix.

In the formulas given below the filters are assumed to be infinitely long and the voltage to be applied at "mid-series" position to the initial or zero-th section. $A_n(t)$ is then equal to the current in the nth section in response to a unit voltage (zero before, unity after time $t=0$.)

1. *Low Pass Wave-Filter, Type L_1C_2, Fig.* 1.

$$A_n(t)=\frac{1}{k}\int_0^x J_{2n}(x)dx, \tag{1a}$$

where $x=\omega_c t$,

$\omega_c=2/\sqrt{L_1C_2}=2\pi$ times the critical or cut-off frequency,

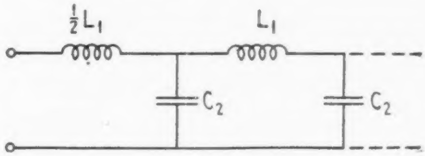


Fig. 1

$J_{2n}(x)$ = The Bessel function of the order $2n$ and argument x, and the filter elements, in terms of the parameters w_c and k, are given by

$$\omega_c = 2/\sqrt{L_1 C_2}, \qquad L_1 = 2k/\omega_c,$$
$$k = \sqrt{L_1/C_2}, \qquad C_2 = 2/\omega_c k.$$

For values of time such that $x<2n$, $A_n(t)$ is very small and positive, while for $x>2n$, the character of the solution is exhibited by the approximate solution

$$A_n(t) = \frac{1}{k}\left[1+\sqrt{\frac{2}{\pi x}}\,\frac{h_{2n}}{q_{2n}}\sin\left(q_{2n}x-\Theta_{2n}\right)\right]. \tag{1b}$$

The formula is deduced from the approximate formulas given in Appendix II for Bessel functions, and h_{2n}, q_{2n} and Θ_{2n} are determined by

$$h_n = \left(\frac{1}{1-n^2/x^2}\right)^{1/4},$$

$$q_n = \sqrt{1-n^2/x^2},$$

and

$$\Theta_n = \frac{2n+1}{4}\pi - n\sin^{-1}(n/x).$$

For sufficiently large values of x, $A_n(t)$ is ultimately given by the asymptotic formula

$$A_n(t) \sim \frac{1}{k}\left[1+\sqrt{\frac{2}{\pi x}}\sin\left(x-\frac{2n+1}{4}\pi\right)\right]. \tag{1c}$$

Formula (1a) was first given by one of the writers (Trans. A. I. E. E., 1919) as a special case of the solution for the dissipative low pass filter (series resistance and shunt leakage).

2. *High Pass Wave-Filter, Type C_1L_2, Fig. 2.*

$$A_n(t) = \frac{1}{k}\left\{\phi_0(x) - \frac{2n}{1!}D^{-1}\phi_1(x) + \frac{(2n)\,(2n-1)}{2!}D^{-2}\phi_2(x) - \ldots\right.$$
$$\left.\ldots + D^{-2n}\phi_{2n}(x)\right\} \tag{2a}$$

where

$x_c = \omega_c t$,

$\omega_c = 2\pi$ times critical frequency *below* which the filter attenuates,

$C_1 = 1/2\omega_c k$, $\quad k = \sqrt{L_2/C_1}$,

$L_2 = k/2\omega_c$, $\quad \omega_c = 1/2\sqrt{L_2 C_1}$.

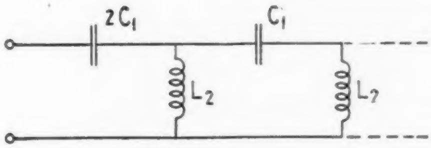


Fig. 2

The symbol D^{-m} denotes multiple integrations, repeated m times and

$$\phi_m(x) = J_o(x) - \frac{m}{1!}J_1(x) + \frac{(m)\,(m-1)}{2!}J_2(x) + \,.\,.\, + (-1)^m J_m(x).$$

A large amount of time and effort have been devoted to an attempt to reduce this and other forms of solution (see Appendix I) to a form in which its properties would be exhibited by direct inspection, but without success. Numerical computations and curves must, therefore, be largely relied upon in the study of the high pass filter in the transient state.

For sufficiently large values of x ($x > 4n^2$) the ultimate behavior of the filter is shown by the asymptotic formula

$$A_n(t) \sim (-1)^n \frac{1}{k}\sqrt{\frac{2}{\pi x}} \cos\,(x - \pi/4). \qquad \text{(2b)}$$

Band Pass Wave-Filters.

In all the band pass types of filters discussed below the transmission band lies in the frequency range between $\omega_1/2\pi$ and $\omega_2/2\pi$ so that the band width is $(\omega_2 - \omega_1)/2\pi$. We shall write $\sqrt{\omega_1\omega_2} = \omega_m$ and $\omega_2 - \omega_1 = w$. For each type the filter elements are determined by the parameters ω_m, w and a third parameter[7] k which may be so chosen as to fix the magnitude of the impedance of the filter.

[7] The parameter k is equal to the characteristic impedance, both mid-series and mid-shunt, at mid-frequency of the confluent band, "constant k" type of wave-filter. See Theory and Design of Uniform and Composite Electric Wave-Filters, this Journal, Jan., 1923.

The formulas for the indicial admittances of all the band pass filters are approximate, as stated above, and are deduced on the assumption that the band width is narrow. Practically, however, as regards the essential deductions drawn from them, they are not so restricted but are applicable to the case of relatively wide bands. (See Appendix I.)

There are, of course, an infinite variety of band pass filters; the ones investigated in the present paper are, however, representative and the conclusions drawn from a study of them are, in their general aspects, applicable to all types.

3. *Band Pass Wave-Filter, Type $L_1C_1L_2C_2$, Fig. 3.*

$$A_n(t)=\frac{w}{\omega_m k}J_{2n}(y)\sin x \tag{3a}$$

where $x=\omega_m t$; $y=wt/2$; and the filter elements are given by

$$L_1=2k/w, \qquad L_2=wk/2\omega_m^2,$$
$$C_1=w/2k\omega_m^2, \qquad C_2=2/wk.$$

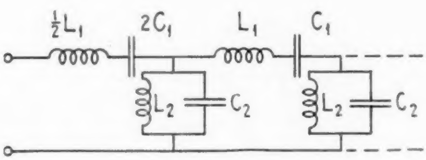


Fig. 3

This is the "constant k" type of filter and, as will be noted, the elements are so proportioned that $L_1C_1=L_2C_2=1/\omega_m^2$, and $L_1/C_2=L_2/C_1=k^2$.

From the properties of Bessel functions discussed in Appendix II, it follows that $A_n(t)$ is very small until $y\geqq 2n$. For values of $y>2n$, the character of the function is clearly exhibited by the following approximate formulas, although these are not sufficiently accurate for the purposes of precise computation.

$$A_n(t)\doteq\frac{w}{\omega_m k}h_{2n}\sqrt{\frac{2}{\pi y}}\cos(q_{2n}y-\Theta_{2n})\sin x \tag{3b}$$

$$\doteq\frac{w}{2\omega_m k}h_{2n}\sqrt{\frac{2}{\pi y}}\left[\sin(x-q_{2n}y+\Theta_{2n})+\sin(x+q_{2n}y-\Theta_{2n})\right] \tag{3c}$$

and ultimately,

$$A_n(t) \sim \frac{w}{2\omega_m k}\sqrt{\frac{2}{\pi y}}\left[\sin\left(x-y+\frac{4n+1}{4}\pi\right)+\sin\left(x+y-\frac{4n+1}{4}\pi\right)\right]. \quad (3d)$$

h_{2n}, q_{2n}, Θ_{2n} are determined by the formulas given in Appendix II,—

$$h_n=\left(\frac{1}{1-n^2/y^2}\right)^{1/4},$$

$$q_n=\sqrt{1-n^2/y^2},$$

and

$$\Theta_n=\frac{2n+1}{4}\pi-n\sin^{-1}(n/y).$$

4. *Band Pass Wave-Filter, Type $L_1C_1C_2$, Fig.* 4.

$$A_n(t)=\frac{w}{\omega_m k}J_n(y)\sin(x-n\pi/2) \quad (4a)$$

where, as above, $x=\omega_m t$; $y=wt/2$, and the filter elements are given by

$$L_1=2k/w;\ C_1=w/2k\omega_1^2;\ C_2=\frac{2}{(\omega_1+\omega_2)k}.$$

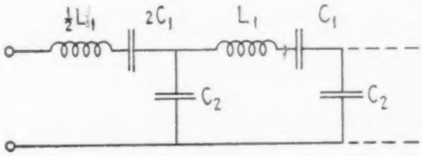


Fig. 4

The approximate formulas for $y>n$, are,

$$A_n(t)\doteq\frac{w}{\omega_m k}h_n\sqrt{\frac{2}{\pi y}}\cos(q_n y-\Theta_n)\sin(x-n\pi/2) \quad (4b)$$

$$\doteq\frac{w}{2\omega_m k}h_n\sqrt{\frac{2}{\pi y}}[\sin(x-q_n y+\Theta_n-n\pi/2)+\sin(x+q_n y-\Theta_n-n\pi/2)] \quad (4c)$$

and ultimately

$$A_n(t)\sim\frac{w}{2\omega_m k}\sqrt{\frac{2}{\pi y}}\left[\sin(x-y+\pi/4)+\sin(x+y-\frac{4n+1}{4}\pi)\right]. \quad (4d)$$

5. *Band Pass Wave-Filter, Type* $L_1C_1L_2$, Fig. 5.

$$A_n(t) = \frac{w}{\omega_m k} J_n(y) \sin(x + n\pi/2) \tag{5a}$$

where $x = \omega_m t$, $y = wt/2$ and the filter elements are determined by

$$L_1 = 2\omega_1 k / w\omega_2; \quad C_1 = w/2k\omega_m^2; \, L_2 = \frac{\omega_1 + \omega_2}{2\omega_m^2} k.$$

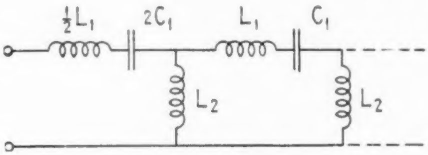


Fig. 5

The approximate formulas for $y > n$ are

$$A_n(t) \doteq \frac{w}{\omega_m k} h_n \sqrt{\frac{2}{\pi y}} \cos(q_n y - \Theta_n) \sin(x + n\pi/2) \tag{5b}$$

$$\doteq \frac{w}{2\omega_m k} h_n \sqrt{\frac{2}{\pi y}} [\sin(x - q_n y + \Theta_n + n\pi/2) + \sin(x + q_n y - \Theta_n + n\pi/2)] \tag{5c}$$

and ultimately

$$A_n(t) \sim \frac{w}{2\omega_m k} \sqrt{\frac{2}{\pi y}} \left[\sin\left(x - y + \frac{4n+1}{4}\pi\right) + \sin(x + y - \pi/4) \right]. \tag{5d}$$

6. *Band Pass Wave-Filter, Type* $L_1L_2C_2$, *Fig.* 6.

$$A_n(t) = \frac{2w}{\omega_m k} [J_n(y) \sin(x - n\pi/2) - J_n'(y) \cos(x - n\pi/2)] \tag{6a}$$

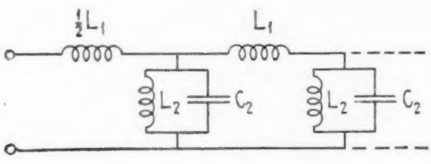


Fig. 6

where $x = \omega_m t$; $y = wt/2$; $J_n'(y) = d/dy\, J_n(y)$, and

$$L_1 = \frac{2k}{\omega_1 + \omega_2}; \; L_2 = wk/2\omega_1^2; \; C_2 = 2/wk.$$

The approximate formulas for $y>n$ are

$$A_n(t) \doteq \frac{2w}{\omega_m k} h_n \sqrt{\frac{2}{\pi y}} [\cos (q_n y - \Theta_n) \sin (x - n\pi/2)$$
$$+ q_n \sin (q_n y - \Theta_n) \cos (x - n\pi/2)] \tag{6b}$$

$$\doteq \frac{w}{\omega_m k} h_n \sqrt{\frac{2}{\pi y}} \begin{bmatrix} (1-q_n) \sin (x - q_n y + \Theta_n - n\pi/2) \\ +(1+q_n) \sin (x + q_n y - \Theta_n - n\pi/2) \end{bmatrix} \tag{6c}$$

and ultimately

$$A_n(t) \sim \frac{2w}{\omega_m k} \sqrt{\frac{2}{\pi y}} \sin\left(x + y - \frac{4n+1}{4}\pi\right). \tag{6d}$$

7. ***Band Pass Wave-Filter, Type $C_1L_2C_2$, Fig. 7.***

$$A_n(t) = \frac{2}{\omega_m k}\left(\frac{w}{2\omega_m}\right)^n P$$
$$+ \frac{2w}{\omega_m k} [J_n(y) \sin (x + n\pi/2) + J_n'(y) \cos (x + n\pi/2)], \tag{7a}$$

where $x = w_m t$; $y = wt/2$, and

$$C_1 = \frac{\omega_1 + \omega_2}{2k\omega_m^2};\ L_2 = wk/2\omega_m^2;\ C_2 = 2\omega_1/w\omega_2 k.$$

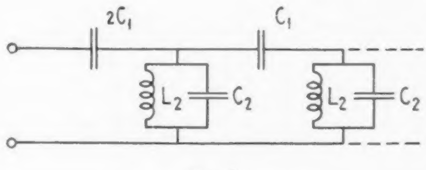


Fig. 7

The symbol P in the first term denotes a "pulse" at time $t=0$; that is

$$P = \infty \text{ at } t = 0,$$
$$= 0 \text{ for } t > 0,$$

and

$$\int_0^\infty P dt = 1.$$

The first term in $A_n(t)$ exists in consequence of the fact that at the instant the voltage is applied the filter behaves like a pure capacity network. For narrow band filters the factor $(w/2\omega_m)^n$ is small so that this term does not contribute appreciably to the steady state. As a matter of fact in actual filters which necessarily have some series resistance, it does not exist.

The approximate formulas for $y>n$ are

$$A_n(t) \doteq \frac{2w}{\omega_m k} h_n \sqrt{\frac{2}{\pi y}} \left[\cos (q_n y-\Theta_n) \sin (x+n\pi/2) - q_n \sin (q_n y-\Theta_n) \cos (x+n\pi/2)\right] \tag{7b}$$

$$\doteq \frac{w}{\omega_m k} h_n \sqrt{\frac{2}{\pi y}} \left[(1+q_n) \sin (x-q_n y+\Theta_n+n\pi/2) + (1-q_n) \sin (x+q_n y-\Theta_n+n\pi/2)\right] \tag{7c}$$

and ultimately

$$A_n(t) \sim \frac{2w}{\omega_m k} \sqrt{\frac{2}{\pi y}} \sin \left(x-y+\frac{4n+1}{4}\pi\right). \tag{7d}$$

8. *Discussion of Indicial Admittances.*

The indicial admittances for the low pass filter, that is, the current in response to a steady unit e.m.f. applied at time $t=0$, are shown in the curves of Figs. 8, 9 and 10, for the initial or zero-th, the 3rd and the 5th sections. These curves together with the exact and approximate formulas given above are sufficient to give a reasonably comprehensive idea of the general character of these oscillations and their dependence on the number of sections and the constants of the filter.

It will be observed that the current is small until a time approximately equal to $2n/\omega_c = n\sqrt{L_1C_2}$ has elapsed after the voltage is applied. Consequently the low pass filter behaves as though currents were transmitted with a finite velocity of propagation $\omega_c/2 = 1/\sqrt{L_1C_2}$ sections per second. This velocity is, however, only apparent or virtual since in every section the currents are actually finite for all values of time>0.

After time $t=n\sqrt{L_1C_2}$ has elapsed the current oscillates about the value $1/k$ with increasing frequency and diminishing amplitude. The amplitude of these oscillations is approximately

$$\frac{1/k}{\sqrt{1-(2n/\omega_c t)^2}} \sqrt{\frac{2}{\pi\omega_c t}}$$

and their instantaneous frequency (measured by intervals between zeros)

$$\frac{\omega_c}{2\pi}\sqrt{1-(2n/\omega_c t)^2}.$$

The oscillations are therefore ultimately of cut-off or critical frequency $w_c/2\pi$ in all sections, but this frequency is approached more and more slowly as the number of filter sections is increased.[8]

The indicial admittances of the band pass filter, type $L_1C_1L_2C_2$, are shown in Figs. 11, 12 and 13 for the initial, the 3rd and the 5th sections. These curves show not the actual oscillations but their *envelopes*. That is to say the curves must be multiplied by sin $\omega_m t$ to give the actual oscillations. The "mid-frequency" $\omega_m/2\pi$ may therefore be regarded as the "carrier frequency" which is modulated by the relatively low frequency oscillations shown in the curves.

Comparison of the formulas for the indicial admittances of the band filters of type $L_1C_1C_2$ and $L_1C_1L_2$ with that of type $L_1C_1L_2C_2$ shows that these curves are applicable to the two former types provided the number of sections is doubled and the phase of the oscillations of frequency $\omega_m/2\pi$ is correctly modified.

Referring to Figs. 11, 12, 13 it will be observed that the oscillations are small until time $t=4n/w$; consequently they are transmitted with an apparent velocity of propagation roughly equal to $w/4=1/2\sqrt{L_1C_2}$[9] sections per second.

After time $t=4n/w$, the low frequency oscillations shown in the curves are of increasing frequency and diminishing amplitude, their envelope being roughly equal to

$$\frac{w}{\omega_m k}\sqrt{\frac{4}{\pi w t}}.$$

The actual oscillations are analyzable into two frequencies

$$\frac{1}{2\pi}\left(\omega_m+\frac{w}{2}\sqrt{1-(4n/wt)^2}\right) \text{ and } \frac{1}{2\pi}\left(\omega_m-\frac{w}{2}\sqrt{1-(4n/wt)^2}\right)$$

so that the ultimate oscillations are of the two critical frequencies

$$\frac{1}{2\pi}(\omega_m+w/2) \text{ and } \frac{1}{2\pi}(\omega_m-w/2).$$

[8] For curves showing the indicial admittance of the low pass filter when n is very large, the reader is referred to Transient Oscillations, Trans. A. I. E. E., 1919.

[9] For types $L_1C_1C_2$ and $L_1C_1L_2$ the velocity in sections per second is double this This corresponds to the fact that two sections of these types are approximately equivalent, as regards their selectivity, to one section of type $L_1C_1L_2C_2$.

For both the low pass and band pass filters the oscillations of the indicial admittances are of continuously variable frequency which traverses the frequency transmission band and ultimately reaches the critical frequencies of the filter.[10]

The indicial admittances of the band pass filter, type $L_1L_2C_2$ are shown in Figs. 14, 15, 16 for the initial, the 6th and 10th sections.[11] The curves show the oscillation envelopes $\sqrt{(J_n^2+J'^2_n)}$, whereas the actual oscillations are within a constant,

$$\sqrt{[J_n^2(wt/2)+J'^2_n(wt/2)]}\ \sin(\omega_m t-n\pi/2-\Theta_n),$$

where $\Theta_n=\tan^{-1}(J_n'/J_n)$. For a narrow band filter the variation in the phase angle Θ_n is very slow.

The principal difference between these curves and the corresponding curves for type $L_1C_1L_2C_2$ is that the envelope of the oscillation does not go through zero as in the latter. In addition the oscillations are ultimately of a single frequency $\frac{1}{2\pi}(\omega_m+w/2)$ while for type $C_1L_2C_2$ the ultimate frequency is $\frac{1}{2\pi}(\omega_m-w/2)$.

The indicial admittance of the high pass filter, shown in the curves of Figs. 17, 18, 19, 20 for the initial, the 1st, 2nd and 3rd sections, differs in important respects from those of the low pass and band pass filters. In the first place the current jumps instantaneously to its maximum value $1/k$ in all sections, so that the velocity of propagation is infinite.[12] After this initial jump the current oscillates with decreasing frequency and decreasing amplitude, the oscillation frequency becoming ultimately the critical or cut-off frequency $\omega_c/2\pi$. The initial frequency and the time required for the oscillation frequency to reduce to $\omega_c/2\pi$, increases, practically linearly with the number of sections. *The oscillation frequency varies continuously and traverses the frequency transmission range of the filter from infinite frequency (represented by the initial jump) down to the critical frequency of the filter, below which it attenuates sinusoidal currents.*

[10] From a purely mathematical viewpoint, this fact explains the transmission, without attenuation, of a continuous band of frequencies.

[11] These curves are applicable to the $C_1L_2C_2$ type of band pass filter, due regard being had to difference in phase, and to the initial jump of current. See formulas (6a) and (7a).

[12] This is, of course, a consequence of the assumption of zero series inductance and shunt capacity. Actually, of course, the circuit must include a finite amount of both.

IV. The Building-Up of Alternating Currents in Wave-Filters

If an e.m.f. $\sin(\omega t+\Theta)$ is applied to the low pass wave-filter (type L_1C_2) at time $t=0$, then by formulas I and (1a), the resultant current in the nth section builds up in accordance with the expression

$$\frac{1}{k}\left[\sin\Theta\int_0^x J_{2n}(x_1)\cos\lambda(x-x_1)dx_1+\cos\Theta\int_0^x J_{2n}(x_1)\sin\lambda(x-x_1)dx_1\right],$$

where $x=\omega_c t$ and $\lambda=\omega/\omega_c$.

For the band pass filter, type $L_1C_1L_2C_2$, the corresponding formula, based on the approximations discussed in the preceding, is by I and (3a).

$$\frac{1}{k}\left[\sin(\mu y+\Theta)\int_0^y J_{2n}(y_1)\cos(\lambda-\mu)(y-y_1)dy_1\right.$$
$$\left.+\cos(\mu y+\Theta)\int^y J_{2n}(y_1)\sin(\lambda-\mu)(y-y_1)dy_1\right]$$

where $y=wt/2$; $\lambda=2\omega/w$; and $\mu=2\omega_m/w$ so that $\mu y=\omega_m t$. Similar formulas are deducible for the other types of band pass filters considered in the preceding section.

Comparison of these formulas shows that, in both the low pass and band pass wave-filters, the genesis and growth of the current in response to an e.m.f. $\sin(\omega t+\Theta)$, applied at time $t=0$, is mathematically determined by definite integrals of the form

$$S_n(z;\nu)=\int_0^z J_n(z_1)\sin\nu(z-z_1)dz_1,$$

and

$$C_n(z;\nu)=\int_0^z J_n(z_1)\cos\nu(z-z_1)dz_1.$$

These integrals [13] have been extensively studied in the course of this investigation; their general properties and the appropriate methods of computation are discussed in Appendix III.

The subsidence of the current, when a sinusoidal e.m.f. is removed, is also determined by the above formulas for the low pass and band pass filters. To show this suppose that prior to the reference time $t=0$, that steady-state currents are flowing in the filter in response

[13] The writers take pleasure in acknowledging their indebtedness to T. H. Gronwall, consulting mathematician, who furnished asymptotic formulas for the computation of these integrals. See Appendix III.

to an e.m.f. sin $(\omega t+\Theta)$, which is removed at time $t=0$. We can represent this condition correctly by regarding the e.m.f. sin $(\omega t+\Theta)$ as continuing, while a negative e.m.f., $-\sin(\omega t+\Theta)$, is applied at time $t=0$. The resultant current for $t\geqq 0$, is then

$$\alpha_n(\omega)\sin(\omega t+\Theta)+\beta_n(\omega)\cos(\omega t+\Theta)$$

$$-\frac{1}{k}\left[\sin\Theta\cdot C_{2n}(x;\lambda)+\cos\Theta\cdot S_{2n}(x;\lambda)\right]$$

for the low pass filter with a corresponding expression for the band pass. $\alpha_n(\omega)$ and $\beta_n(\omega)$ are the real and imaginary parts of the steady state admittances of the filter at frequency $\omega/2\pi$.

Figs. 21–32 exhibit the phenomena attending the building-up of alternating currents in the low pass filter for a sufficient number of representative cases to show the effects of the length of the filter and the applied frequency. For $\omega_c t>25$, the curves represent the ***transient distortion***, that is the difference between the final steady state and actual current. For $\omega_c t<24$ the actual current is shown. An important outstanding result which follows from a study of these curves and the formulas of Appendix III may be stated as follows:

The time T required for an alternating current of frequency $\omega/2\pi$ to build up to its proximate steady state in the nth section of a low pass wave-filter is given approximately by the formula

$$T=\frac{2n}{\omega_c}\frac{1}{\sqrt{1-(\omega/\omega_c)^2}}.$$

The first factor $2n/\omega_c$ represents the delay due to the apparent finite velocity of propagation, while the second factor represents the effect of the applied frequency in its relation to the cut-off frequency of the filter.

This formula is a rather rough approximation when the number of sections n is small. Furthermore the time at which the current reaches its *proximate* steady state does not admit of precise definition.[14] Nevertheless the formula is in substantial agreement with the facts as regards the effect of a number of filter sections, cut-off frequency and applied frequency on the phenomena, and is of great practical importance.[15]

[14] Actually the time T corresponds to a singularity in the mathematical formulas. See Appendix III.

[15] This formula has been applied in the design of periodically loaded cable circuits, which are of such length in the Bell System as to make transient phenomena a factor which must be taken into account. The formula is in close agreement with a large amount of experimental evidence.

The *transient distortion*, it is interesting to note, is, as regards frequency, independent of the applied frequency, and ultimately attains the cut-off frequency of the filter. Its envelope is ultimately

$$\frac{1}{k}\frac{\omega/\omega_c}{1-(\omega/\omega_c)^2}\sqrt{\frac{2}{\pi\omega_c t}}$$

when a voltage $\sin \omega t$ is applied, and

$$\frac{1}{k}\frac{1}{1-(\omega/\omega_c)^2}\sqrt{\frac{2}{\pi\omega_c t}}$$

when a voltage $\cos \omega t$ is applied.

Figs. 33 and 34 show the form of the current in the 5th section when sinusoidal voltages $\sin \omega t$ and $\cos \omega t$ of frequency 25 per cent above the cut-off frequency of the filter are applied. The transient current shown in the curves increases in frequency up to the critical frequency of the filter, the oscillations being ultimately given by

$$\frac{1}{k}\frac{\omega/\omega_c}{(\omega/\omega_c)^2-1}\sqrt{\frac{2}{\pi\omega_c t}}\cos\left(\omega_c t-\frac{2n+1}{4}\pi\right)$$

and

$$\frac{1}{k}\frac{1}{(\omega/\omega_c)^2-1}\sqrt{\frac{2}{\pi\omega_c t}}\sin\left(\omega_c t-\frac{2n+1}{4}\pi\right)$$

corresponding respectively to applied voltages $\sin \omega t$ and $\cos \omega t$. The amplitude of these transient oscillations are enormous compared with the final steady state, and the curves furnish a clear illustration of the fact, stated in a previous part of this paper, that the selective properties of wave-filters are essentially properties of the steady state only.

Figs. 35–41 show the building-up phenomena in the band pass filter, type $L_1C_1L_2C_2$, and are applicable also to types $L_1C_1C_2$ and $L_1C_1L_2$ when proper values are assigned to the constants and parameters.[16] The curves actually show the envelopes of the oscillations which are of slowly variable frequency in the neighborhood of $\omega_m/2\pi$.

A study of these curves and the formulas of Appendix III lead to the following proposition, analogous to that stated above for the low pass filter.

The time T required for an alternating current of frequency $\omega/2\pi$ within the transmission range $w/2\pi$ of a band filter to build up to its

[16] n sections of type $L_1C_1L_2C_2$ are approximately equivalent to $2n$ sections of type $L_1C_1C_2$ or of type $L_1C_1L_2$.

proximate steady state in the nth section is given approximately by the formula

$$T=\frac{4n}{w}\frac{1}{\left[1-4\left(\frac{\omega-\omega_m}{w}\right)^2\right]^{1/2}}$$

for type $L_1C_1L_2C_2$ and one half this amount for the other types of band pass filters discussed in this paper.

These curves show the envelope of the oscillations with fidelity but are not well adapted to exhibit the actual frequencies. These are given by the formula

$$\sqrt{C^2+S^2}\sin\left[\omega_m t+\theta+\tan^{-1}(S/C)\right]$$

where C and S denote the definite integrals

$$C_{2n}\left(\frac{wt}{2};\frac{2(\omega-\omega_m)}{w}\right)\text{ and }S_{2n}\left(\frac{wt}{2};\frac{2(\omega-\omega_m)}{w}\right).$$

The envelope is therefore substantially independent of the phase angle θ of the applied e.m.f. The frequency is ultimately the applied frequency $\omega/2\pi$. The *transient distortion* is analyzable into two frequencies

$$\frac{1}{2\pi}\left(\omega_m+\frac{w}{2}\sqrt{1-(4n/wt)^2}\right)\text{ and }\frac{1}{2\pi}\left(\omega_m-\frac{w}{2}\sqrt{1-(4n/wt)^2}\right),$$

and its envelope is ultimately

$$\frac{1+4\left(\frac{\omega-\omega_m}{w}\right)^2}{1-4\left(\frac{\omega-\omega_m}{w}\right)^2}\sqrt{\frac{4}{\pi wt}},$$

The building-up of alternating currents in the high pass filter has been investigated only qualitatively owing to the extremely laborious computations required. The process is essentially different from that in the low pass and band pass filters. When an e.m.f. $\sin(\omega t+\theta)$ is applied the current in all sections jumps instantly to the value

$$\frac{1}{k}\sin(\omega t+\theta).$$

Therefore the process depends on the applied frequency. If the applied frequency is within the transmission band ($\omega>\omega_c$), the current builds up to its ultimate frequency, the time required being given approximately by the formula

$$T=\frac{2n\omega_c}{\omega^2}\frac{1}{\sqrt{1-\left(\frac{\omega_c}{\omega}\right)^2}},$$

(by the principle of stationary phase; see footnote 31).

It thus requires an infinite time when the applied frequency is equal to the critical frequency while infinite applied frequencies build up instantly.

When the applied frequency is outside the transmission band, the current *subsides* to its steady value, the time required being proportional to the ratio n/ω_c and decreasing as the applied frequency is decreased.

The fact that the initial value of the current is of the same order of magnitude as that of steady state currents in the transmission range is an outstanding feature of the process and reflects the failure of the selective properties of this type of filter in the transient state.

V. The Energy Absorbed from Transient Applied Forces

In only a relatively few cases is the solution for the transient current, in response to suddenly applied forces, reducible to a manageable form, which admits of interpretation or of computation without prohibitive labor. Fortunately, however, it is usually possible to calculate the energy absorbed by a receiving element in a selective network from suddenly applied forces of finite time duration and such a calculation throws a great deal of light on the general properties of selective circuits in the transient state. The calculation is based on formulas VI to IX of Section II.

A particularly important example is the energy absorbed from the force $\sin(pt+\Theta)$, applied at time $t=0$ and removed at time $t=T$. If the energy is averaged with respect to the phase angle Θ, we get [17]

$$\int_0^\infty [I(t)]^2 dt = \frac{1}{2\pi}\int_0^\infty \frac{d\omega}{|Z(i\omega)|^2}\left\{\frac{1-\cos(\omega-p)T}{(\omega-p)^2}+\frac{1-\cos(\omega+p)T}{(\omega+p)^2}\right\}.$$

If $p/2\pi$ is in the neighborhood of the frequency which the network is designed to select, this becomes approximately

$$\int_0^\infty [I(t)]^2 dt = \frac{1}{2\pi}\int_0^\infty \frac{1-\cos(\omega-p)T}{(\omega-p)^2}\,\frac{d\omega}{|Z(i\omega)|^2}. \tag{8}$$

In the *steady state* the time integral of the square of the current in response to the e.m.f. $\sin(pt+\Theta)$ during the time interval T is simply $T/2|Z(ip)|^2$. The expression

$$\frac{|Z(ip)|^2}{\pi T}\int_0^\infty \frac{1-\cos(\omega-p)T}{(\omega-p)^2}\,\frac{d\omega}{|Z(i\omega)|^2} \tag{9}$$

is therefore *the relative amount of energy actually absorbed from the*

[17] Here $Z(i\omega)$ is the steady-state transfer impedance and the integral measures the energy absorbed by a unit resistance in the receiving branch.

force sin $(pt+\theta)$ *acting during the time interval T, to that calculated on the assumption of a steady state in this interval.*

Calculations of these formulas are of particular interest and importance in multiplex carrier telephone and telegraph systems where they furnish a measure of the interference between channels operating at different frequencies.

In order to exhibit clearly the significance of the formulas without detailed computation, consider an ideal selective circuit, for which in the range $\omega_1 \leqq \omega \leqq \omega_2, |Z(i\omega)| = Z_T$ (a constant) and everywhere else $|Z(i\omega)| = Z_s$ (a constant, very large compared with Z_T). Under these assumptions, formulas (8) and (9) become approximately, for the case when $p > \omega_2$,

$$\frac{T}{2Z_s^2} + \frac{1}{2\pi}\frac{\omega_2-\omega_1}{(p-\omega_2)\ (p-\omega_1)}\frac{1}{Z_T^2} \tag{8a}$$

and

$$\left(1+\frac{1}{\pi T}\frac{Z_s^2}{Z_T^2}\frac{\omega_2-\omega_1}{(p-\omega_2)\ (p-\omega_1)}\right). \tag{9a}$$

These formulas admit of some quite interesting deductions which are applicable to band filters in general.

(1) The energy absorbed in excess of that calculated in the steady state basis is

$$\frac{1}{2\pi}\frac{\omega_2-\omega_1}{(p-\omega_2)\ (p-\omega_1)}\frac{1}{Z_T^2}.$$

This is independent of the duration of the applied force and of the degree to which the filter discriminates against steady state currents outside the frequency range $\omega_1 \leqq \omega \leqq \omega_2$. It is proportional to the *band width* and inversely to the product $(p-\omega_2)\ (p-\omega_1)$. It follows therefore that *no amount of selectivity will appreciably reduce the energy absorbed from a sinusoidal force of finite duration outside the transmission range of the filter, below the value given above.*

(2) The fractional excess of energy absorbed is given by

$$\frac{1}{\pi T}\left(\frac{Z_s}{Z_T}\right)^2\frac{\omega_2-\omega_1}{(p-\omega_2)\ (p-\omega_1)}.$$

This decreases with the duration of the applied force but *increases as the square of the selectivity (Z_s/Z_T) of the filter.* Hence for forces of short duration the energy absorbed may be very large compared with that calculated on the steady state basis.

VI. Random Interference

We have hitherto confined attention to the transient phenomena when the form of the applied voltage was explicitly given. In the problem of the behavior of wave-filters and selective circuits in general to such disturbances as "static" in radio transmission and "noise" in wire transmission this is not the case, and the applied force is usually more or less completely *random*. By this it is meant that the interfering disturbance, which may be supposed to originate in a large number of unrelated sources, varies in an irregular, uncontrollable manner, and is characterized statistically by no predominant frequency. Consequently the wave form of the applied force at any particular instant is entirely indeterminate. This fact makes it necessary to treat the problem as a statistical one, and deal with mean values. In the following we shall derive formulas for the mean *energy* absorbed from random interference; and then define and discuss the selective figure of merit of networks with respect to random interference.

The mathematical treatment of the problem will be based on formulas VI to VIII of section II. To apply these formulas to the problem of random disturbances and their effect on selective networks, consider a long interval of time, or epoch, say from 0 to T. During this epoch we suppose that the network is subjected to a large number of individual impressed forces $f_1(t), f_2(t) \ldots f_n(t)$, which are unrelated and vary in intensity and wave form in an irregular, indeterminate manner, and thus constitute what will be called *random interference*. If we write

$$\sum(t) = f_1(t) + f_2(t) + \ldots + f_n(t),$$

then by VI, $\sum(t)$ is representable as a Fourier integral, thus:

$$\sum(t) = \frac{1}{\pi}\int_0^\infty |F(\omega)| \cos[\omega t + \theta(\omega)]\, d\omega$$

while, in accordance with formula VIII, the energy absorbed by the selective network from this random interference is measured by [18]

$$W' = \frac{1}{\pi}\int_0^\infty \frac{|F(\omega)|^2}{|Z(i\omega)|^2}\, d\omega.$$

[18] It should be clearly understood that $Z(i\omega)$ is the transfer impedance of the receiving with respect to the driving branch of the network, and that W' is the energy absorbed by a unit resistance located in the former.

We now introduce the function $R(\omega)$ which will be termed the *energy spectrum of the random interference*, and which is defined by the equation

$$R(\omega)=\frac{1}{T}|F(\omega)|^2. \tag{10}$$

Dividing both sides by T and writing $W'/T=\epsilon$, formula VIII becomes

$$\epsilon=\frac{1}{\pi}\int_0^\infty \frac{R(\omega)}{|Z(i\omega)|^2}d\omega. \tag{11}$$

Both ϵ and $R(\omega)$ become independent of T provided the epoch is made sufficiently great, and ϵ *measures the mean energy absorbed per unit time from the random interference.* The practical significance of this formula is contained in the statement that *the required function of the selective network, as regards random interference, is to minimize the ratio of ϵ to the signal energy. Consequently this ratio furnishes an index of the merit of the network.*

In order to rigorously evaluate the integral of formula (11) the energy spectrum $R(\omega)$ of the interference must be completely specified over the entire interval of integration. Obviously this information cannot be deduced without imposing some restrictions on the character of the interference, or making some hypothesis regarding the mechanism in which it originates. On the other hand if the forces $f_1(t), f_2(t) \ldots f_n(t)$ are absolutely random in a strict mathematical sense, it would appear that all frequencies are equally probable in the spectrum $R(\omega)$ and that, consequently, the most probable energy distribution is that which makes $R(\omega)$ a constant, independent of ω. This inference, however, has not been theoretically established; indeed, the problem does not appear to admit of satisfactory solution by the calculus of probabilities. Furthermore, deductions based on the assumption that the interference is random in a strict mathematical sense might well be inadequate for the applications contemplated, and the "most probable" spectrum in serious disagreement with the spectrum of the actual interference [19] to which we wish to apply the results of the present study.

Fortunately, in view of these difficulties, a complete specification of $R(\omega)$ is not at all necessary for a practical solution of the problem. This is a consequence of the following facts:

[19] For example, the spectrum of the interference presented to the terminals of the selective network will be modified by the characteristics of the "transducer," over which the disturbances are transmitted. Thus both in radio and wire systems, the greater attenuation suffered in transmission by high frequencies, will reduce the relative intensity of the high frequency part of the spectrum.

(a) In the case of efficient selective networks, the important contributions to the integral (11) are confined to a finite continuous range of ω which includes, but is not greatly in excess of, the range which the network is designed to select.[20] This fact is a consequence of the impedance characteristics of selective networks and of the following properties of the spectrum $R(\omega)$.

(b) $R(\omega)$ is a continuous, finite function of ω which converges to zero at infinity and is everywhere positive. It possesses no sharp maxima or minima,[21] and its variation with respect to ω, where it exists, is slow. These properties of $R(\omega)$ are believed to be evident from physical considerations, and will not be elaborated.

Now referring to formula (11), since the numerator and denominator of the integrand are everywhere positive, it follows that a value ω_m of ω exists, such that

$$\epsilon = \frac{1}{\pi} R(\omega_m) \int_0^\infty \frac{d\omega}{|Z(i\omega)|^2}.$$

Now suppose that the network is designed to select frequencies in the range $\omega_1 \leqq \omega \leqq \omega_2$. Then from the properties of the network and of the spectrum $R(\omega)$ discussed above, it follows that ω_m lies close to, or within, the range $\omega_1 \leqq \omega \leqq \omega_2$. In any case, if the band $\omega_2 - \omega_1$ is made so narrow that the curvature of $R(\omega)$ over the interval is negligible, then with negligible error ω_m may be taken as 2π times the "mid-frequency" of the band. That is to say, with negligible error, ω_m may be defined either as $(\omega_1+\omega_2)/2$ or as $\sqrt{\omega_1\omega_2}$.

The foregoing argument may be summarized in the following proposition:

The mean energy ϵ absorbed per unit time from random interference by a selective network designed to select the band of frequencies corresponding to $\omega_1 \leqq \omega \leqq \omega_2$ is measured by the formula

$$\epsilon = \frac{1}{\pi} \rho R(\omega_m), \tag{12}$$

where ρ denotes the infinite integral

$$\rho = \int_0^\infty \frac{d\omega}{|Z(i\omega)|^2}.$$

[20] This statement excludes from present consideration networks, which, like the high pass filter, select an infinite band of frequencies. This limitation, however, is of no practical consequence, because such networks are quite useless as regards random interference. This question will be briefly discussed later.

[21] The existence of sharp maxima and minima would indicate the presence of *systematic* interference, which should not be regarded as part of the random interference.

and $R(\omega_m)$ *is the spectral energy level of the interference at frequency* $\omega_m/2\pi$. ω_m *lies close to or within the band* $\omega_1<\omega<\omega_2$, *and when this band is sufficiently small with respect to the curvature of* $R(\omega)$, $\omega_m/2\pi$ *may be taken as the mid-frequency of the band.*

Formula (12) is of very considerable practical and theoretical importance. It furnishes a basis for the experimental determination of the energy spectrum $R(\omega)$, and this determination, for any given epoch, can be made as accurate as desired by employing a band filter which selects a sufficiently narrow band of frequencies. It also leads immediately to the following important proposition.

If a selective network is required to select the band of frequencies corresponding to $\omega_1 \leqq \omega \leqq \omega_2$, *the mean energy absorbed per unit time by the network from random interference is necessarily greater than*

$$\frac{1}{\pi}\int_{\omega_1}^{\omega_2}\frac{R(\omega)}{|Z(i\omega)|^2}d\omega \doteq \frac{1}{\pi}R(\omega_m)\int_{\omega_1}^{\omega_2}\frac{d\omega}{|Z(i\omega)|^2}. \tag{13}$$

This formula, therefore, determines the theoretical limit, beyond which it is not possible to discriminate against random interference.

We are now prepared to introduce a formula which defines *the figure of merit of a selective network with respect to random interference.* This formula gives the signal-to-random-interference energy ratio of the network as compared with the corresponding ratio in an ideal reference circuit (defined below).

Let the network, as above, be designed to select frequencies in the band $\omega_1 \leqq \omega \leqq \omega_2$. Then the energy absorbed per unit time from *steady-state* forces in this frequency range is proportional to

$$\sigma = \frac{1}{\omega_2-\omega_1}\int_{\omega_1}^{\omega_2}\frac{d\omega}{|Z(i\omega)|^2}.$$

The corresponding mean energy absorbed from random interference is proportional to

$$\rho = \int_0^\infty \frac{d\omega}{|Z(i\omega)|^2}$$

when the energy level of the interference is corrected to unity.

The ratio $S=\sigma/\rho$ defines the selective figure of merit of the network with respect to random interference.

Stated in words, *the selective figure of merit of a network with respect to random interference is equal to the statistical signal-to-random-interference energy ratio, divided by the corresponding ratio in an ideal band filter which transmits without loss all frequencies in a "unit" band* $(\omega_2-\omega_1=1)$, *and absolutely extinguishes all frequencies outside this band.*

In the foregoing argument, the theoretical limitations have been carefully pointed out and even emphasized. In practical applications, however, it is believed that these limitations are of small or negligible importance, and that the formula for and definition of the selective figure of merit furnish all the information, as regards the behavior of selective circuits to random interference, which we are in a position to make use of. Thus the formula is immediately applicable to the problem of determining the effect of band width, number of sections, dissipation, and terminal reflections on the selectivity of filters with respect to random interference. It furnishes likewise, a means of estimating the comparative merits of the very large number of circuits which have been invented for the purpose of eliminating "static" in radio communication, and leads to general deductions of practical value regarding the inherent limitations imposed on the solution of the "static" problem.

The utility and significance of the foregoing formulas will now be illustrated by application to some representative selective circuits. It is easily shown that, to a good approximation, in the case of the low pass filter (type L_1C_2)

$$S = \frac{1}{\omega_c(1+1/16n^2)},$$

and for the band pass filter (type $L_1C_1L_2C_2$)

$$S = \frac{1}{w(1+1/16n^2)}.$$

In these formulas n denotes the number of filter sections while ω_c is 2π times the cut-off frequency of the low pass filter and w is 2π times the transmission band width of the band filter. In both cases the filters are assumed to be terminated in their characteristic impedances and to be non-dissipative.[22] These formulas show at once the effect of band width and number of sections n on the behavior of wave-filters to random interference, and lead to the following proposition.

In filters designed to select a band of frequencies of width w, the ratio of energy transmitted through the network by the signal and by random interference is inversely proportional to the band width and increased inappreciably when the number of sections is increased beyond two.

As regards the effect of dissipation, a second proposition is deducible.

The effect of introducing dissipation into a network designed to select a single frequency or a band of frequencies is always such as to reduce the ratio of signal energy to that absorbed from random interference.

[22] These approximate formulas are in very good agreement with actual calculations for filters terminated in resistances.

An inference drawn from the study of band filters in the preceding section may be stated as follows:

The selective figure of merit of a wave-filter designed to select a finite band of frequencies is approximately proportional to the minimum time required for sinusoidal currents within the transmission band to build up their approximate steady values, divided by the number of filter sections.

Another circuit of practical interest, which has been proposed as a solution of the "static" problem in radio-communication consists of a series of sharply tuned oscillation circuits, unilaterally coupled through amplifiers.[23] This circuit is designed to receive only a single frequency to which all the individual oscillation circuits are tuned. The figure of merit of this circuit is approximately

$$S=\frac{L}{R}\,\frac{2^{2n-2}[(n-1)!]^2}{(2n-2)!}$$

where n denotes the number of sections, or stages, and L and R are the inductance and resistance of the individual oscillation circuits. The outstanding fact in this formula is the slow rate of increase of S with the number of stages. For example, if the number of stages is increased from 1 to 5, the figure of merit increases only by the factor 3.66, while for a further increase in n the gain is very slow. This gain, furthermore, is accompanied by a serious increase in the "sluggishness" of the circuit; that is, in the particular example cited, by an increase of 5 to 1 in the time required for signals to build up to their steady-state.[24]

The outstanding deduction of practical importance to be drawn from the preceding is that, as regards disturbances which are predominantly random, irregular, or discontinuous, it is useless to employ selective circuits of extremely high selectivity. The gain in signal-to-interference ratio is very small when the selectivity is increased beyond a moderate amount, and is only gotten by making the circuit relatively sluggish and slowly responsive.

The preceding discussion is, for the reasons discussed above, not applicable to selective circuits like the high pass filter, which transmit an infinite band of frequencies. Considerable information, however, regarding the behavior of the high pass filter to random disturbances can be gotten by returning to formula (10) and comparing the energy absorbed by the high pass filter, with that absorbed by a *pure-resistance network*. Reference to formula (10) shows that the

[23] See U. S. Patent No. 1173079 to Alexanderson.

[24] When the number of stages n is fairly large, the selective figure of merit becomes proportional to $\sqrt{n}$ and the building-up time to n.

energy absorbed from random disturbances by a pure resistance network is proportional to

$$\int_0^\infty R(\omega)d\omega.$$

The relative amount of energy absorbed by the high pass filter is *greater than*

$$\int^\infty R(\omega)d\omega.$$

The function $R(\omega)$ represents, as above, the statistical energy spectrum of the interference.

Comparison of these formulas shows at once that, unless the energy of the random interference is largely confined in the range $\omega < \omega_c$, little protection is afforded by the high pass filter.

Appendix I

Derivation of Wave-Filter Indicial Admittances

1. *Low Pass Wave-Filter, Type L_1C_2.*

The derivation of the indicial admittance of this type of filter is given in detail by one of the writers in a previous paper.[25] The method of solution there employed, which is quite generally applicable to periodic structures, consists in writing down the Heaviside Expansion formula for the current in the nth section of a filter of s sections in length $(s>n)$, short circuited at the sth section. The expansion is converted into a definite integral by letting s become infinite and the formula becomes that of the indicial admittance of the nth section of an infinitely long filter. For the non-dissipative filter having mid-series termination, this procedure leads to the formula

$$A_n(t)=\frac{1}{k}\int_0^x dx_1\frac{2}{\pi}\int_0^{\pi/2}\cos{(2n\lambda)}\cdot\cos{(x_1\sin\lambda)}d\lambda, \qquad x=\omega_c t,$$

which is identifiable, from known formulas, as

$$A_n(t)=\frac{1}{k}\int_0^x J_{2n}(x_1)dx_1. \tag{1.1}$$

A much more direct and flexible method of solution and one which avoids the necessity of setting up the Heaviside expansion formula

[25] Transient Oscillations, Trans. A. I. E. E., 1919. This paper should be consulted for the details of this method.

and then converting into a definite integral, is to employ the integral equation II. If $Z_n(p)$ denote the transfer operational impedance of the nth section of the infinitely long low pass filter, we have

$$\frac{1}{Z_n(p)}=\frac{\omega_c}{k}\frac{1}{\sqrt{p^2+\omega_c^2}}\left(\frac{\sqrt{p^2+\omega_c^2}-p}{\omega_c}\right)^{2n} \tag{1.2}$$

and writing $x=\omega_c t$, $F_n(x)=kA_n(t)$, the integral equation II becomes

$$\int_0^\infty F_n(x)e^{-px}dx=\frac{1}{p\sqrt{p^2+1}}\left(\sqrt{p^2+1}-p\right)^{2n}. \tag{1.3}$$

The solution of this integral equation is known[26]; it is

$$F_n(x)=\int_0^x J_{2n}(x_1)dx_1$$

which agrees with the preceding.

The "mid-series" termination is chosen not only for its importance in practical applications but because in general the indicial admittance has been found to take the simplest form when the voltage is applied at this position. This is not always the case, however. For example in the low pass filter if the e.m.f. is applied, not directly at mid-series but through a terminal inductance $L=L_1/2=k/\omega_c$, the integral equation becomes

$$\int_0^\infty F_n(x)e^{-px}dx=\left(1-p(\sqrt{p^2+1}-p)\right)\frac{1}{p\sqrt{p^2+1}}\left(\sqrt{p^2+1}-p\right)^{2n},$$

whence

$$F_n(x)=\int_0^x J_{2n}(x_1)dx_1-J_{2n+1}(x). \tag{1.4}$$

Unless, however, the terminal impedance is related in some simple manner to the constants of the filter, the resulting formula is necessarily complicated.

2. *High Pass Wave-Filter, Type C_1L_2.*

For this type of filter it can be shown, by the first method discussed above in connection with the low pass filter, that the indicial admittance is expressible as the definite integral

$$A_n(t)=\frac{2}{\pi k}\int_1^\infty\frac{\cos\left(2n\sin^{-1}\frac{1}{\lambda}\right)\sin x\lambda\, d\lambda}{\sqrt{\lambda^2-1}}, \tag{2.1}$$

[26] Nielsen, Cylinderfunktionen, page 186, formula 13.

where $x=\omega_c t$. For the case $n=0$, the solution can be recognized as

$$A_0(t)=\frac{1}{k}J_0(x).$$

To attack the problem by means of the integral equation II, we write down the operational transfer impedance

$$\frac{1}{Z_n(p)}=\frac{1}{k}\,\frac{1}{\sqrt{1+\omega_c^2/p^2}}\left(\sqrt{1+\omega_c^2/p^2}-\omega_c/p\right)^{2n}. \tag{2.2}$$

Writing $\omega_c t=x$, and $A_n(t)=\frac{1}{k}F_n(x)$, and substituting in II gives, as the integral equation of the problem

$$\int_0^\infty F_n(x)e^{-px}dx=\frac{1}{p^{2n}}\,\frac{1}{\sqrt{p^2+1}}\left(\sqrt{p^2+1}-1\right)^{2n}. \tag{2.3}$$

The solution of this equation can be expressed in a number of different forms, depending on the type of expansion of the right hand side which we adopt. One form is as follows:

Expansion of the bracketted expression on the right hand side by the binomial theorem and rearrangement gives

$$\int_0^\infty F_n(x)e^{-px}dx=\frac{1}{\sqrt{p^2+1}}\left[\frac{1}{p^{2n}}+\frac{(2n)(2n-1)}{2!}\,\frac{(p^2+1)}{p^{2n}}+\frac{(2n)\,.\,.\,(2n-3)}{4!}\right.$$
$$\left.\times\frac{(p^2+1)^4}{p^{2n}}+\,.\,.\right]-\left[\frac{2n}{1!}\,\frac{1}{p^{2n}}+\frac{(2n)(2n-1)(2n-2)}{3!}\,\frac{(p^2+1)}{p^{2n}}+\,.\,.\,.\,.\,.\right].$$

Recognizing that

$$\int_0^\infty J_0(x)e^{-px}dx=\frac{1}{\sqrt{p^2+1}},$$

the solution, after rearrangement, becomes the terminating series [27]

$$F_n(x)=kA_n(t)$$
$$=\left(1+\frac{4n^2}{2!}D^{-2}+\frac{4n^2(4n^2-2^2)}{4!}D^{-4}+\frac{4n^2(4n^2-2^2)(4n^2-4^2)}{6!}D^{-6}+\,.\,.\right)J_0(x)$$
$$-2n\left(x+\frac{(4n^2-2^2)}{3!}\,\frac{x^3}{3!}+\frac{(4n^2-2^2)(4n^2-4^2)}{5!}\,\frac{x^5}{5!}+\,.\,.\,.\right), \tag{2.4}$$

where D^{-m} indicates multiple integration, repeated m times. Thus

$$D^{-1}J_0(x)=\int_0^x J_0(x_1)dx_1;\quad D^{-2}J_0(x)=\int_0^x dx_1\int_0^{x_1}J_0(x_2)dx_2;\text{ etc.}$$

[27] This solution has been derived from the definite integral also.

Another type of expansion, leading to the formula given in the text, is suggested by the known identity

$$\int_0^\infty J_n(x)e^{-px}dx = \frac{1}{\sqrt{p^2+1}}(\sqrt{p^2+1}-p)^n.$$

To introduce this identity, we write the integral equation in the form

$$\int_0^\infty F_n(x)e^{-px}dx = \frac{1}{p^{2n}}\frac{1}{\sqrt{p^2+1}}\Big((\sqrt{p^2+1}-p)+(p-1)\Big)^{2n}$$

and expand the bracketted expression by the binomial theorem. Identification of the individual terms and rearrangement gives the terminating series

$$F_n(x) = \phi_0(x) - \frac{2n}{1!}D^{-1}\phi_1(x) + \frac{(2n)(2n-1)}{2!}D^{-2}\phi_2(x) - \ldots$$

$$- \frac{2n}{1!}D^{-(2n-1)}\phi_{2n-1}(x) + D^{-2n}\phi_{2n}(x),$$

where $\phi_m(x)$ denotes the terminating series

$$\phi_m(x) = J_0(x) - \frac{m}{1!}J_1(x) + \frac{(m)(m-1)}{2!}J_2(x) + \ldots + (-1)^m J_m(x)$$

and as above D^{-m} denotes multiple integration.

It is an easy matter to derive solutions in the form of infinite series, as for example power series and Bessel series. These solutions, however, which have been carefully investigated, have not proved manageable for either computation or interpretation. The solutions given above are also unfortunately, extremely difficult to compute or interpret. For computation, numerical integration of the following difference equations, is sometimes preferable

$$F_0(x) = J_0(x),$$

$$F_1(x) - F_0(x) = 2\int_0^x dx_1 \int_0^{x_1} F_0(x_2)dx_2 - 2x,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots \tag{2.5}$$

$$F_{n+1}(x) - 2F_n(x) + F_{n-1}(x) = 4\int_0^x dx_1 \int_0^{x_1} F_n(x_2)dx_2,$$

$$n \geqq 1.$$

3. *Band Pass Wave-Filter.*

The mathematical discussion of the band pass filters will be limited to the $L_1C_1L_2C_2$ type shown in Fig. 3. This type is representative

and the appropriate mathematical procedure is essentially the same for all the band pass wave-filters.

The first method of solution outlined above for the low pass and high pass filters, leads, for the $L_1C_1L_2C_2$ type of band pass filter, to the definite integral formula

$$A_n(t) = \frac{w}{\omega_m k}\frac{2}{\pi}\int_0^{\pi/2} \frac{\sin gx}{g} \cos 2n\mu \cdot \cos(y \sin \mu) d\mu, \qquad (3.1)$$

where $x = \omega_m t$; $y = wt/2$; $\rho = w/2\omega_m$; and

$$g = \sqrt{1+\rho^2 \sin^2 \mu}.$$

In solving this definite integral, use is made of the known formulas,

$$J_{2n}(y) = \frac{2}{\pi}\int_0^{\pi/2} \cos 2n\mu \cdot \cos(y \sin \mu) d\mu \qquad (3.2)$$

and

$$(-1)^s \frac{d^{2s}}{dy^{2s}} J_{2n}(y) = \frac{2}{\pi}\int_0^{\pi/2} \sin^{2s}\mu \cdot \cos 2n\mu \cdot \cos(y \sin \mu) d\mu. \qquad (3.3)$$

If in (3.1) g is replaced by unity, it follows from (3.2) that, to this approximation

$$A_n(t) = \frac{w}{\omega_m k} J_{2n}(y) \sin x \qquad (3.4)$$

which is formula (3a) of the text [28]. Clearly this becomes an increasingly good approximation as the parameter ρ becomes smaller; that is, as the ratio of the band width $w/2\pi$ to the mid-frequency $\omega_m/2\pi$ becomes smaller. The approximate formulas of the text for the other types of band pass filters were derived by precisely similar procedure and involve approximations of the same character and order of magnitude.

To investigate the approximate solution, we proceed as follows: If we write

$$\frac{\omega_m k}{w} A_n(t) = F_n(x,y) = \frac{2}{\pi}\int_0^{\pi/2} \frac{\sin gx}{g} \cos 2n\mu \cdot \cos(y \sin \mu) d\mu, \qquad (3.5)$$

and

$$G_n(x,y) = \frac{2}{\pi}\int_0^{\pi/2} \sin gx \cdot \cos 2n\mu \cdot \cos(y \sin \mu) d\mu, \qquad (3.6)$$

and if we substitute for $1/g$ in (3.5) the expansion

$$1 - \frac{1}{2}\rho^2 \sin^2\mu + \frac{1 \cdot 3}{2 \cdot 4}\rho^4 \sin^4\mu - \ldots\ldots,$$

[28] If a series resistance R_1 and a shunt resistance $R_2 = k^2/R_1$ are included in the filter sections, the formula becomes (3.4) multiplied by the factor $\exp(-R_1 y/2k)$.

it follows from a formula exactly analagous to (3.3) that

$$F_n(x,y) = \left(1 + \frac{1}{2}\rho^2 \frac{\partial^2}{\partial y^2} + \frac{1\cdot 3}{2\cdot 4}\rho^4 \frac{\partial^4}{\partial y^4} + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\rho^6 \frac{\partial^6}{\partial y^6} + \ldots\right) G_n(x,y) \quad (3.7)$$

so that the problem is reduced to the solution of the definite integral $G_n(x, y)$.

In the integral (3.6), write $g = 1 + h$, so that

$$h = \sqrt{1 + \rho^2 \sin^2\mu} - 1, \quad (3.8)$$

whence

$$G_n(x,y) = \sin x \cdot \frac{2}{\pi}\int_0^{\pi/2} \cos hx \cdot \cos 2n\mu \cdot \cos(y \sin \mu) d\mu$$
$$+ \cos x \cdot \frac{2}{\pi}\int_0^{\pi/2} \sin hx \cdot \cos 2n\mu \cdot \cos(y \sin \mu) d\mu \quad (3.9)$$
$$= P_n \sin x + Q_n \cos x, \quad (3.10)$$

where P_n and Q_n denote the definite integrals of (3.9). This effects a further reduction of the problem to the solution of the definite integrals P_n and Q_n.

In the integrands of these integrals expand $\cos hx$ and $\sin hx$ in the usual power series, and in each term thereof introduce the expansion

$$h^s = \left(\frac{\rho^2}{2}\right)^s (\sin^{2s}\mu)\,(1 + a_{s1}\rho^2 \sin^2\mu + a_{s2}\rho^4 \sin^4\mu + \ldots),$$

where the coefficients are given by

$$a_{sj} = (-1)^j s \frac{(s+2j-1)!}{(s+j)!\; j!}\left(\frac{1}{4}\right)^j.$$

By aid of this procedure it is easily shown that

$$P_n = J_{2n}(y) - \frac{(\rho^2 x/2)^2}{2!}\frac{d^4}{dy^4}\left(1 - a_{21}\rho^2 \frac{d^2}{dy^2} + a_{22}\rho^4 \frac{d^4}{dy^4} - \ldots\right) J_{2n}(y)$$
$$+ \frac{(\rho^2 x/2)^4}{4!}\frac{d^8}{dy^8}\left(1 - a_{41}\rho^2 \frac{d^2}{dy^2} + a_{42}\rho^4 \frac{d^4}{dy^4} - \ldots\right) J_{2n}(y)$$
$$- \frac{(\rho^2 x/2)^6}{6!}\frac{d^{12}}{dy^{12}}\left(1 - a_{61}\rho^2 \frac{d^2}{dy^2} + a_{62}\rho^4 \frac{d^4}{dy^4} - \ldots\right) J_{2n}(y)$$
$$+ \ldots, \quad (3.11)$$

with a corresponding expansion formula for Q_n.

It is now convenient to introduce the symbolic notation

$$P_n = \cos\left[x(\sqrt{1-\rho^2 d^2}-1)\right] J_{2n}(y) \tag{3.12}$$

and

$$Q_n = \sin\left[x(\sqrt{1-\rho^2 d^2}-1)\right] J_{2n}(y) \tag{3.13}$$

where the symbol d denotes the differential operator d/dy operating on $J_{2n}(y)$. The actual numerical significance of these formulas is gotten by expanding as in (3.11).

With the same symbolic notation we get finally,

$$A_n(t) = \frac{w}{\omega_m k} \sin\left(x\sqrt{1-\rho^2 d^2}\right) \frac{1}{\sqrt{1-\rho^2 d^2}} J_{2n}(y). \tag{3.14}$$

The exact solution (3.14) is too complicated, as it stands, to be of any practical value. Fortunately, however, it is possible to sum the expression asymptotically, and the resultant formula shows clearly the behavior of $A_n(t)$ and in particular the character and magnitude of the errors in the approximate formula of the text.

When y is large compared with $(4n)^2$,

$$J_{2n}(y) \doteq \sqrt{\frac{2}{\pi y}} \cos\left(y - \frac{4n+1}{4}\pi\right)$$

and (3.15)

$$\frac{d^{2s}}{dy^{2s}} J_{2n}(y) \doteq (-1)^s \sqrt{\frac{2}{\pi y}} \left\{ \begin{aligned} &\left[1 - \frac{3}{2}\frac{2s(2s-1)}{4y^2}\right] \cos\left(y - \frac{4n+1}{4}\pi\right) \\ &- \frac{s}{y} \sin\left(y - \frac{4n+1}{4}\pi\right) \end{aligned} \right.$$

to order $1/y^2$.

If this expression is substituted in the expanded form of (3.14), some rather intricate and tedious operations finally give as the asymptotic limit of $A_n(t)$

$$A_n(t) \sim \frac{w}{\omega_m k} \sqrt{\frac{2}{\pi y}} \left\{ \begin{aligned} &\left(1 - \frac{1}{8}\rho^2 + \ldots\right) \sin\left(x\sqrt{1+\rho^2}\right) \cos\left(y - \frac{4n+1}{4}\pi\right) \\ &- \left(\frac{1}{2}\rho + \ldots\right) \cos\left(x\sqrt{1+\rho^2}\right) \sin\left(y - \frac{4n+1}{4}\pi\right). \end{aligned} \right. \tag{3.16}$$

The coefficients of the two terms of (3.16) are even and odd power series in ρ respectively, powers of ρ beyond the second being neglected.

Formula (3.16) is important, as showing the effect of the band width, that is of the parameter ρ, on the indicial admittance. It can be used for numerical computation, however, only when $y > (4n)^2$. A corresponding formula, valid over a much wider range, is obtain-

able from the expression derived in Appendix II for the Bessel function, namely

$$J_{2n}(y) = B_{2n}(y) \cos \Omega_{2n}(y).$$

If this expression is employed instead of (3.15), we get corresponding to (3.16),

$$A_n(t) \doteq \frac{w}{\omega_m k} B_{2n}(y) \left\{ \begin{array}{l} \left(1-\frac{1}{8}\sigma^2+\ldots\right) \sin (x\sqrt{1+\sigma^2}) \cos \Omega_{2n}(y) \\ -\left(\frac{1}{2}\sigma+\ldots\right) \cos (x\sqrt{1+\sigma^2}) \sin \Omega_{2n}(y) \end{array} \right. \quad (3.17)$$

$$\doteq \frac{w}{\omega_m k} \left\{ \begin{array}{l} \left(1-\frac{1}{8}\sigma^2+\ldots\right) \sin (x\sqrt{1+\sigma^2})\ J_{2n}(y) \\ +\left(\frac{1}{2}\rho+\ldots\right) \cos (x\sqrt{1+\sigma^2})\ J'_{2n}(y), \end{array} \right. \quad (3.18)$$

where $\sigma = \rho q_{2n} = \rho\sqrt{1-(2n/y)^2}$.

Formula (3.17) is valid when $y > 2n$, and ultimately approaches the limit (3.16) as y becomes indefinitely large.

We are now prepared to discuss the character of the approximations of the formula of the text, which may be written as

$$\frac{w}{2\omega_m k} B_{2n}(y) \left\{ \sin [x+\Omega_{2n}(y)] + \sin [x-\Omega_{2n}(y)] \right\}. \quad (3.19)$$

Correspondingly (3.17) may be written as

$$\frac{w}{2\omega_m k} B_{2n}(y) \left\{ \begin{array}{l} \left(1-\frac{1}{2}\sigma+\ldots\right) \sin [x\sqrt{1+\sigma^2}+\Omega_{2n}(y)] \\ +\left(1+\frac{1}{2}\sigma+\ldots\right) \sin [x\sqrt{1+\sigma^2}-\Omega_{2n}(y)]. \end{array} \right. \quad (3.20)$$

Comparison of (3.19) and (3.20) shows that the approximate formula of the text ignores slowly variable correction factors in the amplitudes of the component oscillations, and a slowly variable change in their frequencies. For band pass filters employed in practice these corrections are not only slowly variable but in most cases are quite small. In any case, it is important to observe that failure to include these corrections does not appreciably affect any essential features of the building-up phenomena discussed in the text. Consequently the deductions from the formula of the text are valid not only for narrow-band pass filters, but also for filters of quite wide bands. This statement is substantiated by the fact that the steady-state characteristics,

deduced from the approximate formula in accordance with the general formula V, are in excellent agreement with the exact values.

As illustrating the appropriate methods in the solution of problems in electric circuit theory, it is of interest to derive the formula for the band pass filter directly from the integral equation II. The method is not only more generally applicable, but avoids the necessity of deriving the definite integral (3.1). We therefore start with the formulas:

$$\int_0^\infty e^{-pt} A_n(t)dt = 1/p\ Z_n(p)$$

or

$$\int_0^\infty e^{-pt} A'_n(t)dt = 1/Z_n(p), \text{ where } A'_n(t) = d/dt\ A_n(t).$$

For all wave-filters of the "ladder" type it may be shown that

$$\frac{1}{Z_n(p)} = \frac{1}{z_2} \frac{(\sqrt{1+r/4} - \sqrt{r/4})^{2n}}{\sqrt{r+r^2/4}}, \tag{3.21}$$

where z_1 and z_2 are the series and shunt impedances respectively, and $r = z_1/z_2$. This expression admits of series expansion

$$\begin{aligned}\frac{1}{Z_n(p)} = \frac{2}{z_1}\Bigg[\frac{1}{r^n} - \frac{2n+2}{1!}\frac{1}{r^{n+1}} + \frac{(2n+3)\ (2n+4)}{2!}\frac{1}{r^{n+2}} \\ - \frac{(2n+4)\ (2n+5)\ (2n+6)}{3!}\frac{1}{r^{n+3}} + \ldots\Bigg].\end{aligned} \tag{3.22}$$

For the $L_1C_1L_2C_2$ type of filter

$$1/r = \left(\frac{w}{2}\right)^2 \left(\frac{p}{p^2+\omega_m^2}\right)^2$$

and

$$1/z_1 = \frac{1}{k}\left(\frac{w}{2}\right)\left(\frac{p}{p^2+\omega_m^2}\right).$$

It follows from (3.22) and the integral identity,

$$\int_0^\infty e^{-pt} A'_n(t)dt = 1/Z_n(p)$$

that $A'_n(t)$ has an expansion solution of the form

$$\begin{aligned}A'_n(t) = \frac{w}{k}\Bigg\{ \rho^{2n} f_{2n}(x) - \frac{2n+2}{1!} \rho^{2n+2} f_{2n+2}(x) + \\ + \frac{(2n+3)\ (2n+4)}{2!} \rho^{2n+4} f_{2n+4}(x) \\ - \frac{(2n+4)\ (2n+5)\ (2n+6)}{3!} \rho^{2n+6} f_{2n+6}(x) \ldots\ldots \Bigg\},\end{aligned} \tag{3.23}$$

where $x=\omega_m t$; $\rho=w/2\omega_m$; and the $f_s(x)$ functions are defined and determined by the integral identities,

$$\int_0^\infty f_s(x)e^{-px}dx=\left(\frac{p}{p^2+1}\right)^{s+1} \tag{3.24}$$

for all integral values of s.

For $s=0$, the solution of this equation is known; it is

$$f_0(x)=\cos x.$$

The solutions for $s>0$ are gotten from the recurrence formulas [29]

$$f_s(x)=\int_0^x \cos(x-\lambda)\,f_{s-1}(\lambda)d\lambda.$$

Repeated applications of this formula give

$$f_{2s}(x)=\frac{1}{2^{2s}}\left(P_{2s}(x)\cos x+Q_{2s}(x)\sin x\right)$$

where P_{2s} and Q_{2s} are polynomials in x of the $2s^{th}$ and $(2s-1)^{th}$ orders respectively. Thus:

$$P_{2s}=\alpha(s)\frac{x^{2s}}{2s!}+\beta(s)\frac{x^{2s-2}}{(2s-2)!}+\gamma(s)\frac{x^{2s-4}}{(2s-4)!}+\ .\ .$$

(terminating in term in $x^2/2!$),

and

$$Q_{2s}=a(s)\frac{x^{2s-1}}{(2s-1)!}+b(s)\frac{x^{2s-3}}{(2s-3)!}+c(s)\frac{x^{2s-5}}{(2s-5)!}+\ .\ .$$

(terminating in term in $x/1!$).

The α, β, γ . . . a, b, c, . . . coefficients are functions of the order s; the first few coefficients are:

$$\alpha(s)=1,\ s\geqq 0,$$

$$a(s)=\frac{2s+1}{2},\ s\geqq 1,$$

$$\beta(s)=-\frac{(2s-2)(2s+1)}{8},\ s\geqq 2,$$

$$b(s)=\frac{(2s-2)(2s+1)}{8}-\frac{(2s-3)(2s+1)(2s+2)}{3\cdot 16},\ s\geqq 2.$$

If the foregoing expressions for the f_s functions are substituted in the series solution (3.23) for $A_n'(t)$ and if the series are rearranged as explained below, we get writing $wt/2=\rho x=y$,

$$A_n'(t)=\frac{w}{k}\left\{J_{2n}(y)\cos x+\rho R_1(y)\sin x+\rho^2 R_2(y)\cos x\ .\ .\right\}.$$

[29] See equation 10, The Heaviside Operational Calculus, *B. S. T. J.*, Nov., 1922.

The first term $J_{2n}(y)$ is gotten by picking out the leading terms in the P polynomials; the second term $R_1(y)$ by picking out the leading terms in the Q polynomials; the third term $R_2(y)$, from the second terms in the P polynomials; etc.

The work of rearranging and identifying the "remainder" functions $R_1(y)$, $R_2(y)$ . . . is rather intricate and tedious. The first few functions can be written as

$$R_1(y) = \left(\frac{y}{2}\right)\left(\frac{d^2}{dy^2} + \frac{2}{y}\frac{d}{dy}\right)J_{2n}(y),$$

$$R_2(y) = -\frac{1}{2!}\left(\frac{y}{2}\right)^2\left(\frac{d^4}{dy^4} + \frac{4}{y}\frac{d^3}{dy^3}\right)J_{2n}(y),$$

$$R_3(y) = -\frac{1}{3!}\left(\frac{y}{2}\right)^3\left(\frac{d^6}{dy^6} + \frac{6}{y}\frac{d^5}{dy^5} - \frac{6}{y^2}\frac{d^4}{dy^4} - \frac{24}{y^3}\frac{d^3}{dy^3}\right)J_{2n}(y), \text{ etc.}$$

If we substitute these expressions, rearrange and write $\rho y/2 = z$, we get finally

$$A_n'(t) = \frac{w}{k}\left\{\begin{array}{l} \cos x\left[1 - \dfrac{z^2}{2!}\dfrac{d^4}{dy^4} + \dfrac{z^4}{4!}\dfrac{d^8}{dy^8} \dots\right]J_{2n}(y) \\ + \sin x\left[\dfrac{z}{1!}\dfrac{d^2}{dy^2} - \dfrac{z^3}{3!}\dfrac{d^6}{dy^6} + \dots\right]J_{2n}(y) \\ + \rho \sin x\left[\dfrac{d}{dy} - \dfrac{z^2}{2!}\dfrac{d^5}{dy^5} + \dots\right]J_{2n}(y) \\ - \rho \cos x\left[\dfrac{z}{1!}\dfrac{d^3}{dy^3} - \dfrac{z^3}{3!}\dfrac{d^7}{dy^7} + \dots\right]J_{2n}(y) \\ + \text{series involving factors in } \rho^2 \text{ and higher powers.} \end{array}\right.$$

Neglecting factors in ρ^2, this becomes

$$A_n(t) = \frac{w}{w_m k}\left\{\begin{array}{l} \sin x\left[1 - \dfrac{z^2}{2!}\dfrac{d^4}{dy^4} + \dfrac{z^4}{4!}\dfrac{d^8}{dy^8} \dots\right]J_{2n}(y) \\ - \cos x\left[\dfrac{z}{1!}\dfrac{d^2}{dy^2} - \dfrac{z^3}{3!}\dfrac{d^6}{dy^6} + \dfrac{z^5}{5!}\dfrac{d^{10}}{dy^{10}} \dots\right]J_{2n}(y). \end{array}\right.$$

The character of this solution in the region $y > 2n$, is shown by the asymptotic approximation

$$A_n(t) = \frac{w}{\omega_m k} J_{2n}(y) \sin\left(1 + \tfrac{1}{2}\rho^2 q_{2n}^2\right)x \qquad (3.25)$$

where

$$q_{2n} = \sqrt{1 - \frac{(2n)^2}{y^2}}.$$

To the same order of approximation in $\rho = w/2\omega_m$, this agrees with the solution (3.18) given above.

Appendix II

Properties of the Bessel Function $J_n(x)$

The Bessel functions have been studied and tabulated more exhaustively than any other functions largely owing to their great importance and frequent occurrence in mathematical physics. Qualitatively their behavior for integral orders n and real arguments x may be described as follows.

When the argument is less than the order $(0 \leqq x < n)$ the function is very small and positive, and is initially zero (except when $n = 0$). In the neighborhood of $x = n$, the function begins to build up and reaches a maximum a little beyond the point $x = n$. Thereafter the function oscillates with increasing frequency and diminishing amplitude, and ultimately behaves as

$$\sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{2n+1}{4}\pi\right).$$

When $n = 0$, the initial value is unity, but the subsequent behavior of the function is as described above.

In order to get a more accurate picture of this function the following approximate formula was developed in the course of the present investigation.[30]

$$J_n(x) \doteq B_n(x) \cos \Omega_n(x), \qquad \text{for } x > n$$

where

$$B_n(x) = \sqrt{\frac{2}{\pi x}} \frac{1}{\left(1 - \frac{m^2}{x^2} + \frac{3}{2}\frac{m^2}{x^4}\frac{1}{(1 - m^2/x^2)^2}\right)^{1/4}},$$

$$\Omega_n(x) = x\left[\sqrt{1 - \frac{m^2}{x^2}} + \frac{m}{x}\sin^{-1}\left(\frac{m}{x}\right) - \frac{m^2}{4x^4}\frac{1}{(1 - m^2/x^2)^{3/2}}\right] - \frac{2n+1}{4}\pi,$$

$$\Omega_n'(x) = \frac{d}{dx}\Omega_n(x),$$

$$= \sqrt{1 - \frac{m^2}{x^2} + \frac{3}{2}\frac{m^2}{x^4}\frac{1}{(1 - m^2/x^2)^2}},$$

and

$$m^2 = n^2 - 1/4.$$

[30] It was subsequently discovered that somewhat similar formulas had previously been developed by Graf and Gubler (Einleitung in die Theorie der Besselschen Funktionen), and by Nicholson (*Phil. Mag.*, 1910, p. 249).

This approximate formula is valid only where $x>n$, its accuracy increasing with x and with n. For all orders of n it is quite accurate beyond the first zero of the function.

The "instantaneous frequency" of oscillation is approximately

$$\frac{1}{2\pi}\Omega_n'(x)=\frac{1}{2\pi}\sqrt{1-\frac{m^2}{x^2}+\frac{3}{2}\frac{m^2}{x^4}\frac{1}{(1-m^2/x^2)^2}}.$$

By this it is meant that at any point x $(>n)$ the interval between successive zeros is approximately $\pi/\Omega'(x)$. Otherwise stated, in the neighborhood of any point x, the function behaves like a sinusoid of amplitude $B_n(x)$ and frequency $\omega/2\pi$ where $\omega=\Omega_n'(x)$.

The following approximate formulas, while not sufficiently precise for the purposes of accurate computation except for quite large values of x, clearly exhibit the character of the functions for values of the argument $x>n$, and of the order $n>2$.

$$J_n(x)\doteq h_n\sqrt{\frac{2}{\pi x}}\cos(q_nx-\Theta_n),$$

$$J_n'(x)=-q_nh_n\sqrt{\frac{2}{\pi x}}\sin(q_nx-\Theta_n),$$

$$\int_0^x J_n(x)dx=1+\frac{h_n}{q}\sqrt{\frac{2}{\pi x}}\sin(q_nx-\Theta_n),$$

where

$$h_n=\left(\frac{1}{1-n^2/x^2}\right)^{1/4}\doteq 1+\frac{n^2}{4x^2},$$

$$q_n=\sqrt{1-n^2/x^2},$$

and

$$\Theta_n=\frac{2n+1}{4}\pi-n\sin^{-1}(n/x).$$

Appendix III

Building-Up of Alternating Currents in Wave-Filters

The integrals

$$S_n(z;\nu)=\int_0^z J_n(z_1)\sin\nu(z-z_1)dz_1$$

and

$$C_n(z;\nu)=\int_0^z J_n(z_1)\cos\nu(z-z_1)dz_1,$$

on which the genesis and growth of alternating currents in the low pass and band pass filters depends, have been computed as follows.

For values of $z<24$, $n\leqq 10$ and $\nu\leqq 1$, they are accurately calculable from the following series expansions

$$C_n(z;\nu)=2(c_1J_{n+1}(z)+c_3J_{n+3}(z)+c_5J_{n+5}(z)+\ \cdot\ \cdot),$$

and

$$S_n(z;\nu)=4\nu(c_2J_{n+2}(z)+c_4J_{n+4}(z)+c_6J_{n+6}(z)+\ \cdot\ \cdot),$$

where the coefficients c_1, c_2 . . . are polynomials in 2ν, and are independent of the index n. They are

$$c_1=1,$$

$$c_3=1-(2\nu)^2,$$

$$c_5=1-\frac{3}{1!}(2\nu)^2+(2\nu)^4,$$

$$c_7=1-\frac{3\cdot 4}{2!}(2\nu)^2+\frac{5}{1!}(2\nu)^4-(2\nu)^6,$$

$$c_9=1-\frac{3\cdot 4\cdot 5}{3!}(2\nu)^2+\frac{5\cdot 6}{2!}(2\nu)^4-\frac{7}{2!}(2\nu)^6+(2\nu)^8,$$

. . . .

$$c_2=1,$$

$$c_4=\frac{2}{1!}-(2\nu)^2,$$

$$c_6=\frac{2\cdot 3}{2!}-\frac{4}{1!}(2\nu)^2+(2\nu)^4,$$

$$c_8=\frac{2\cdot 3\cdot 4}{3!}-\frac{4\cdot 5}{2!}(2\nu)^2+\frac{6}{1!}(2\nu)^4-(2\nu)^6,$$

. . . .

The tabulation of $J_n(z)$ for values of z up to 24 and of n up to 60 given by Gray and Mathews and by Jahnke und Emde make the computation for integral values of z rapid and precise.

For large values of n the integrals can be accurately computed, except in the neighborhood of the critical point $z=n/\sqrt{1-\nu^2}$, $(\nu<1)$, from the asymptotic formulas furnished by Gronwall.

Without detailed computation, however, the general character of the integrals can be shown as follows with an accuracy usually sufficient for engineering purposes. By differentiation S_n and C_n satisfy the differential equations

$$S'_n=\nu C_n,$$

and

$$C'_n=J_n(z)-\nu S_n,$$

where the primes denote differentiation with respect to the argument z. The solution of these differential equations is based on the approximation, valid only when $z>n$,

$$\frac{d^2}{dz^2}J_n(z) \doteq -q_n^2 J_n(z), \qquad q_n=\sqrt{1-n^2/z^2}.$$

To this approximation, which becomes more and more accurate as z and n increase, the differential equations are satisfied by solutions of the form

$$S_n=\frac{\nu}{\nu^2-q_n^2}J_n(z)+A\sin(\nu z-\alpha),$$

and

$$C_n=\frac{1}{\nu^2-q_n^2}J_n'(z)+A\cos(\nu z-\alpha).$$

A and α in the complementary terms are arbitrary constants, which must be determined. These complementary terms, periodic in νz, are evidently the ultimate values of the integrals when z approaches infinity, which are known. Other considerations, however, show that these terms should be omitted when $\nu<1$ and $z<n/\sqrt{1-\nu^2}$. Consequently we arrive at the following approximations.[31]

For $\nu<1$ *and* $n<z<n/\sqrt{1-\nu^2}$,

$$S_n(z;\nu)=\frac{\nu}{\nu^2-q_n^2}J_n(z),$$

$$C_n(z;\nu)=\frac{1}{\nu^2-q_n^2}J_n'(z),$$

and

$$q_n=\sqrt{1-n^2/z^2}.$$

This approximation is not accurate at $z=n$, and breaks down at the critical point $z=n/\sqrt{1-\nu^2}$. In the interval between, however, it is a fair approximation, particularly when ν is nearly equal to unity and n is not too small.

For $\nu<1$ *and* $z>n/\sqrt{1-\nu^2}$,

$$S_n(z;\nu)=\frac{\nu}{\nu^2-q_n^2}J_n(z)+\frac{1}{\sqrt{1-\nu^2}}\sin(\nu z-n\sin^{-1}\nu),$$

and

$$C_n(z;\nu)=\frac{1}{\nu^2-q_n^2}J_n'(z)+\frac{1}{\sqrt{1-\nu^2}}\cos(\nu z-n\sin^{-1}\nu).$$

[31] The qualitative properties of these definite integrals can be deduced from the principle of stationary phase (See Theory of Bessel Functions, G. N. Watson, p. 229).

This formula can be safely employed only when z considerably exceeds the critical value $n/\sqrt{1-\nu^2}$.

For $\nu>1$ *and* $z>n$, the ultimate periodic terms are very small, and may be omitted unless n is too small. Consequently in this region,

$$S_n(z;\nu)\doteq\frac{\nu}{\nu^2-q_n^2}J_n(z),$$

and

$$C_n(z;\nu)\doteq\frac{1}{\nu^2-q_n^2}J'_n(z).$$

In the range of values for which the foregoing approximations are valid we have also to the same approximation (see Appendix II)

$$J_n(z)\doteq\sqrt{\frac{2}{\pi z}}\cos\,(q_nz-\Theta_n),$$

and

$$J'_n(z)\doteq-q_n\sqrt{\frac{2}{\pi z}}\sin\,(q_nz-\Theta_n).$$

Appendix IV

The Effects of Terminal Impedances

In the text of this paper, the calculation of the wave-filter indicial admittances is based on the assumption that the voltage is applied directly to the filter at "mid-series" position and that the filter is either infinitely long or else, what amounts to the same thing, is terminated in its characteristic impedance. By virtue of these assumptions, the disturbing effects of terminal reflections are eliminated, and, as shown in the text, the solution is reducible to a relatively simple form, which admits of considerable instructive interpretation by inspection, and is rather easily computed.

In the following the general solution will be given for the indicial admittance $A_n(t)$ in the nth section of a wave-filter of s sections or length, with the e.m.f. applied to the initial or zero-th section through an impedance $Z_1(p)=Z_1$ and the last or sth section closed by an impedance $Z_2(p)=Z_2$.

For any type of periodic structure, including as a limiting case, the smooth line, it can readily be shown that

$$\frac{1}{Z_n(p)}=\sigma\frac{1}{K_1}\,\frac{e^{-n\Gamma}+\rho_2e^{-(2s-n)\Gamma}}{1-\rho_1\rho_2e^{-2s\Gamma}}\tag{1}$$

where

K_1 = characteristic impedance, as seen from terminals of initial or zero-th section,

K_2 = characteristic impedance, as seen from terminals of last or sth section,

Γ = propagation constant per section,

Z_1, Z_2 = terminal impedances,

$$\sigma=\frac{K_1}{K_1+Z_1},$$

$$\rho_1=\frac{K_1-Z_1}{K_1+Z_1},$$

and

$$\rho_2=\frac{K_2-Z_2}{K_2+Z_2}.$$

K_1, K_2, Z_1, Z_2, and consequently σ, ρ_1, ρ_2 are, of course, functions of the operator p.

The corresponding indicial admittance $A_n(t)$ is given by the integral equation

$$\int_0^\infty e^{-pt}A_n(t)dt=\frac{1}{pZ_n(p)}. \tag{2}$$

By aid of (1) the right hand side of (2) can be expanded as

$$\sigma\frac{e^{-n\Gamma}}{pK_1}+\sigma\rho_2\frac{e^{-(2s-n)\Gamma}}{pK_1}+\sigma\rho_1\rho_2\frac{e^{-(2s+n)\Gamma}}{pK_1}+\sigma\rho_1\rho_2^2\frac{e^{-(4s-n)\Gamma}}{pK_1}$$
$$+\sigma\rho_1^2\rho_2^2\frac{e^{-(4s+n)\Gamma}}{pK_1}+\ .\ .\ .\ .\ .\ .\ . \tag{3}$$

Now if $a_m(t)$ denotes the indicial admittance in the mth section of an infinitely long periodic structure, when the e.m.f. is applied directly to the sending end terminals, it follows from (2) and (3) that

$$\int_0^\infty e^{-pt}a_m(t)dt=\frac{e^{-m\Gamma}}{pK_1}. \tag{4}$$

From (2), (3) and (4) it follows at once that

$$A_n(t) = \frac{d}{dt}\int_0^t dy \left\{ \begin{array}{l} r_0(t-y)a_n(y) + r_1(t-y)a_{2s-n}(y) + r_2(t-y)a_{2s+n}(y) \\ + r_3(t-y)a_{4s-n}(y) + r_4(t-y)a_{4s+n}(y) + \dots \end{array} \right. \quad (5)$$

provided the functions $r_0(t)$, $r_1(t)$, $r_2(t)$ . . . satisfy, and are defined by, the equations

$$\int_0^\infty e^{-pt} r_0(t)dt = \frac{\sigma}{p} = \frac{1}{p}\frac{K_1}{K_1 + Z_1},$$

$$\int_0^\infty e^{-pt} r_1(t)dt = \frac{\sigma\rho_2}{p} = \frac{1}{p}\frac{K_1}{K_1 + Z_1}\cdot\frac{K_2 - Z_2}{K_2 + Z_2}, \quad (6)$$

$$\int_0^\infty e^{-pt} r_2(t)dt = \frac{\sigma\rho_1\rho_2}{p} = \frac{1}{p}\frac{K_1}{K_1 + Z_1}\cdot\frac{K_1 - Z_1}{K_1 + Z_1}\cdot\frac{K_2 - Z_2}{K_2 + Z_2}, \text{ etc.}$$

If the indicial admittance in any section of an infinitely long periodic structure is determined, and equations (6) solved for $r_0(t), r_1(t), r_2(t)$ . . . (by aid of any of the methods discussed in the present paper), then $A_n(t)$ is given by (5) by a single quadrature. The solution may appear quite involved; as a matter of fact it is the simplest and most easily interpreted and computed form of solution possible and its complexity merely reflects the complicated character of reflection effects due to terminal impedances. This considered statement is made in the light of an extensive study of the whole problem and the literature bearing on it and has been tested in many specific cases.

When the terminal impedances Z_1 and Z_2 are complicated and entirely unrelated to the corresponding characteristic impedances K_1 and K_2, the solution of equations (6) and the numerical computations of (5) are laborious but entirely possible, the only questions being as to whether the importance of the problem justifies the necessary expenditure of time and effort. In many cases, also, approximate solutions are obtainable. Without any computations, however, the solution (5) admits of considerable instructive interpretation by inspection. The first term represents the current in the nth section of an infinitely long structure when a unit e.m.f. is impressed through a terminal impedance Z_1. $r_0(t)$ is the corresponding voltage which exists across the terminals proper. The second term is a reflected wave from the other terminals due to the terminal impedance irregularity which exists there. The third term is a reflected wave from the sending end terminals due to the corresponding terminal impedance irregularity, etc. The solution, consequently, is expanded in a form which corresponds exactly with the actual sequence of phenomena which occur.

The solution takes a particularly simple and instructive form when $Z_1 = k_1 K_1$ and $Z_2 = k_2 K_2$ where k_1 and k_2 are numerics. In this case the solutions of (6) give

$$r_0(t) = r_0 = \frac{1}{1+k_1},$$

$$r_1 = \frac{1}{1+k_1} \cdot \frac{1-k_2}{1+k_2},$$

$$r_2 = \frac{1}{1+k_1} \cdot \frac{1-k_1}{1+k_1} \cdot \frac{1-k_2}{1+k_2}, \quad \text{etc. and}$$

$$A_n(t) = \frac{1}{1+k_1} \left\{ a_n(t) + \frac{1-k_2}{1+k_2} a_{2s-n}(t) + \frac{1-k_1}{1+k_1} \cdot \frac{1-k_2}{1+k_2} a_{2s+n}(t) + \ldots \right\}.$$

The solution for the special cases of open and short circuit terminations follow at once by assigning the values of zero or infinity, as the case may be, to k_1 and k_2. If $k_1 = o$; $k_2 = 1$, $A_n(t)$ reduces to $a_n(t)$ as, of course, it should.

BOOK B
TIGHT

BOUND

HTLY

Low Pass Wave-Filter, Type L_1C_2

Divide ordinates by k and abscissae by ω_c to read current in amperes and time in seconds.

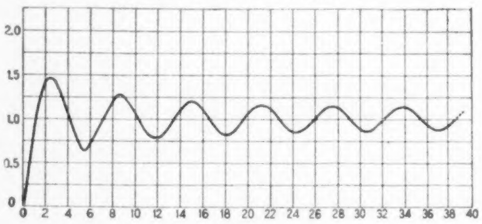


Fig. 8
Indicial Admittance of Initial Section.

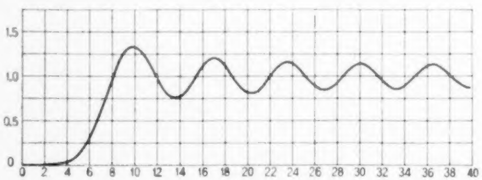


Fig. 9
Indicial Admittance of Third Section.

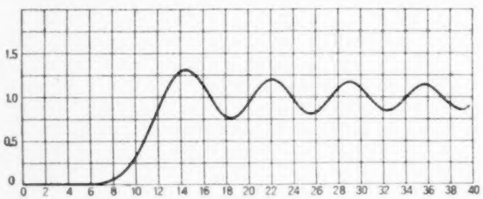


Fig. 10
Indicial Admittance of Fifth Section.

Band Pass Wave-Filter, Type $L_1C_1L_2C_2$

Divide ordinates by $\omega_m k/w$ and abscissae by $w/2$ to read current in amperes and time in seconds.

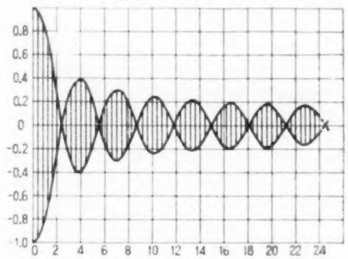


Fig. 11
Indicial Admittance of Initial Section.

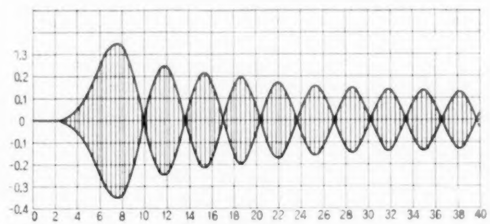


Fig. 12
Indicial Admittance of Third Section.

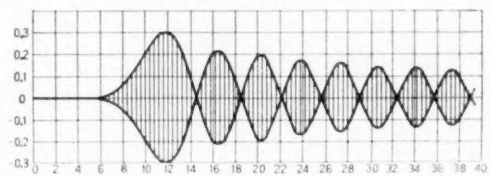


Fig. 13
Indicial Admittance of Fifth Section.

Band Pass Wave-Filter, Type $L_1L_2C_2$

Divide ordinates by $\omega_m k/2w$ and abscissae by $w/2$ to read current in amperes and time in seconds.

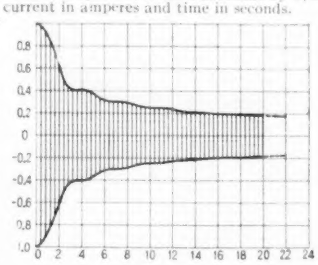


Fig. 14
Indicial Admittance of Initial Section.

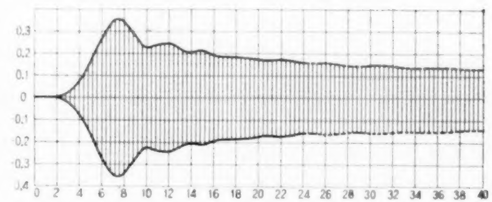


Fig. 15
Indicial Admittance of Sixth Section.

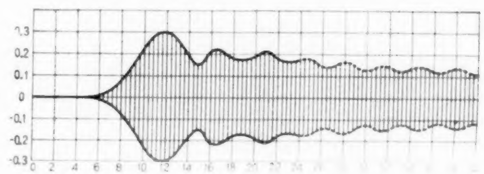


Fig. 16
Indicial Admittance of Tenth Section.

High Pass Wave-Filter, Type C_1L_2

Divide ordinates by k and abscissae by ω_c to read current in amperes and time in seconds.

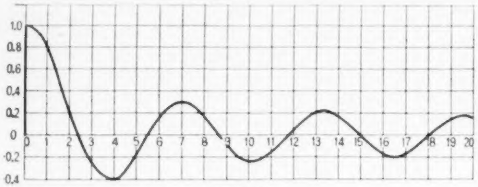


Fig. 17
Indicial Admittance of Initial Section.

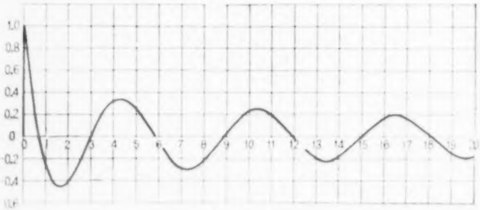


Fig. 18
Indicial Admittance of First Section.

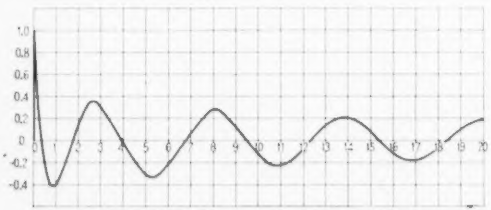


Fig. 19
Indicial Admittance of Second Section.

High Pass Wave-Filter, Type C_1L_2—*Contd.*

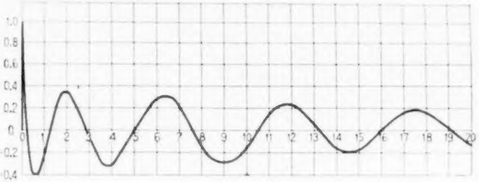


Fig. 20
Indicial Admittance of Third Section.

Low Pass Wave-Filter, Type L_1C_2

Divide ordinates by k and abscissae by ω_c to read current in amperes and time in seconds.

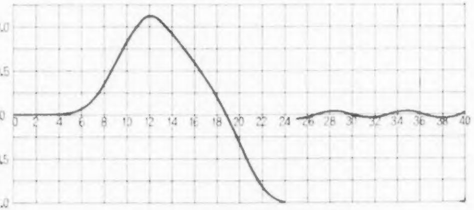


Fig. 21
Current in Third Section in Response to E.M.F. sin $\frac{1}{4}\omega_c t$.

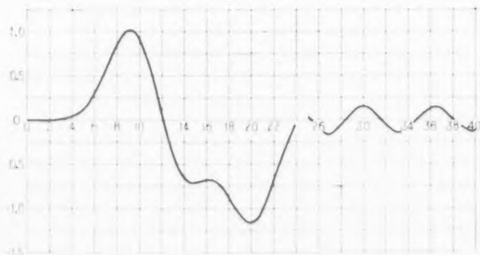


Fig. 22
Current in Third Section in Response to E.M.F. cos $\frac{1}{4}\omega_c t$.

Low Pass Wave-Filter, Type L_1C_2—*Contd.*

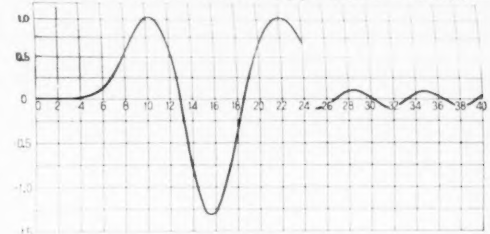


Fig. 23
Current in Third Section in Response to E.M.F. sin $\frac{1}{2}\omega_c t$.

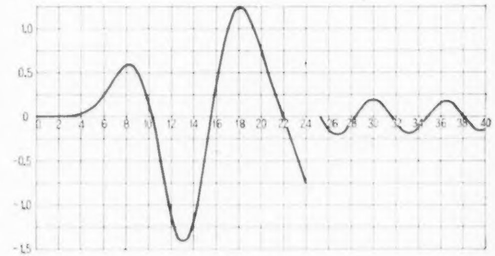


Fig. 24
Current in Third Section in Response to E.M.F. cos $\frac{1}{2}\omega_c t$.

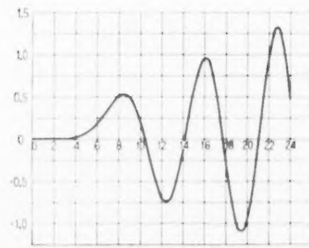


Fig. 25
Current in Third Section in Response to E.M.F. sin $\omega_c t$.

Low Pass Wave-Filter, Type L_1C_2—*Contd.*

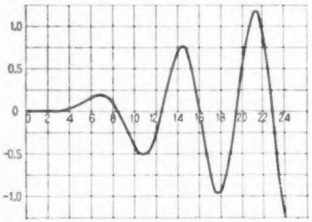


Fig. 26
Current in Third Section in Response to E.M.F. $\cos \omega_c t$.

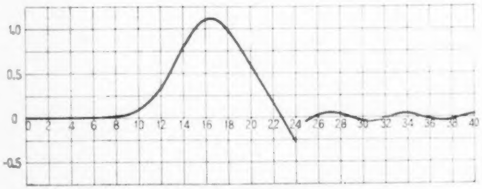
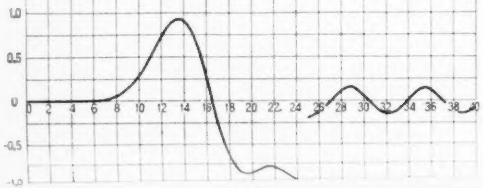


Fig. 27
Current in Fifth Section in Response to E.M.F. $\sin \frac{1}{2}\omega_c t$.



Low Pass Wave-Filter, Type L_1C_2—*Contd.*

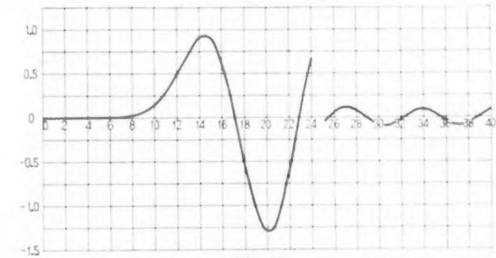


Fig. 29
Current in Fifth Section in Response to E.M.F. $\sin \frac{1}{2}\omega_c t$.

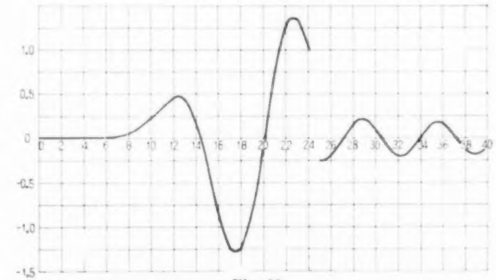
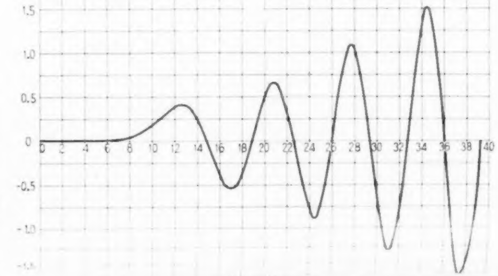


Fig. 30
Current in Fifth Section in Response to E.M.F. $\cos \frac{1}{2}\omega_c t$.



Low Pass Wave-Filter, Type L_1C_2—*Contd.*

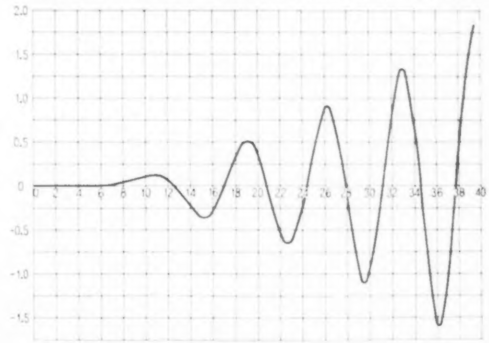


Fig. 32
Current in Fifth Section in Response to E.M.F. $\cos \omega_c t$.

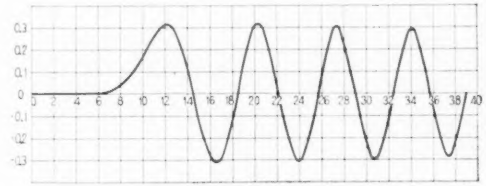
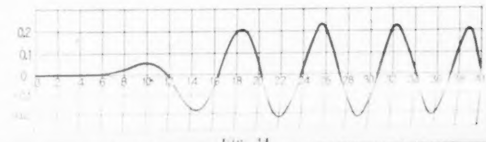


Fig. 33
Current in Fifth Section in Response to E.M.F. $\sin 1.25\, \omega_c t$.
Steady-state Amplitude = 0.0013.



Band Pass Wave-Filter, Type $L_1C_1L_2C_2$
Divide ordinates by ω_{mk}/w and abscissae by $w/2$ to read

Band Pass Wave-Filter, Type $L_1C_1L_2C_2$, *Contd.*

Band Pass Wave-Filter, Type $L_1C_1L_2C_2$, *Contd.*

Band Pass Wave-Filter, Type $L_1C_1L_2C_2$

Divide ordinates by $\omega_m k/w$ and abscissae by $w/2$ to read current in amperes and time in seconds.

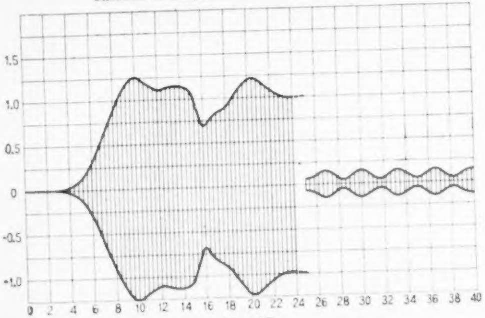


Fig. 35

Envelope of Current in Third Section in Response to E.M.F. of Frequency $\frac{1}{2\pi}(\omega_m \pm w/8)$.

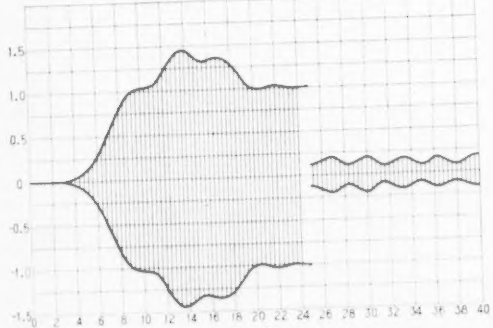


Fig. 36

Envelope of Current in Third Section in Response to E.M.F. of Frequency $\frac{1}{2\pi}(\omega_m \pm w/4)$.

Band Pass Wave-Filter, Type $L_1C_1L_2C_2$, *Contd.*

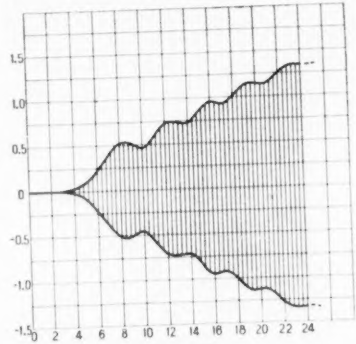


Fig. 37

Envelope of Current in Third Section in Response to E.M.F. of Frequency $\frac{1}{2\pi}(\omega_m \pm w/2)$.

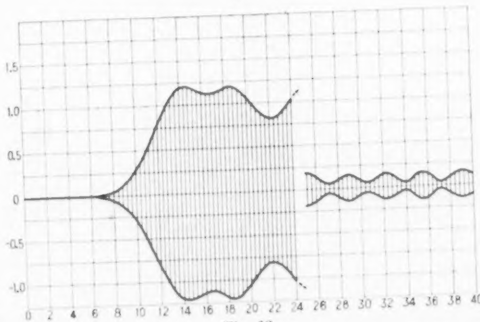


Fig. 38

Envelope of Current in Fifth Section in Response to E.M.F. of Frequency $\frac{1}{2\pi}(\omega_m \pm w/8)$.

Band Pass Wave-Filter, Type $L_1C_1L_2C_2$, *Contd.*

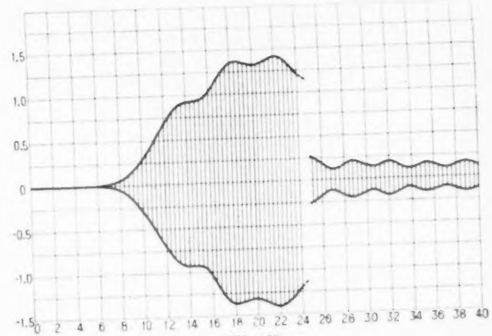


Fig. 39

Envelope of Current in Fifth Section in Response to E.M.F. of Frequency $\frac{1}{2\pi}(\omega_m \pm w/4)$.

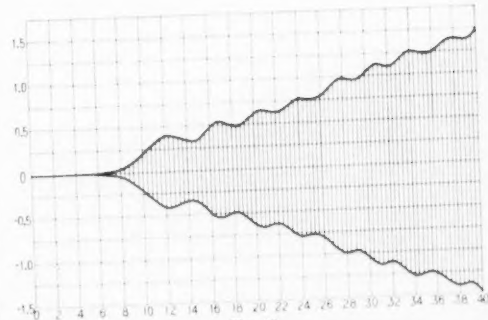


Fig. 40

Envelope of Current in Fifth Section in Response to E.M.F. of Frequency $\frac{1}{2\pi}(\omega_m \pm w/2)$.

Band Pass Wave-Filter, Type $L_1C_1L_2C_2$, *Contd.*

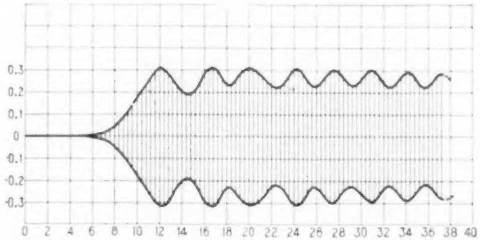


Fig. 41

Envelope of Current in Fifth Section in Response to E.M.F. of Frequency $\frac{1}{2\pi}\left(\omega_m \pm \frac{5}{4}w/2\right)$.

Band Pass Wave-Filter, Type $L_1L_2C_2$

Divide ordinates by $\omega_m k/w$ and abscissae by $w/2$ to read current in amperes and time in seconds.

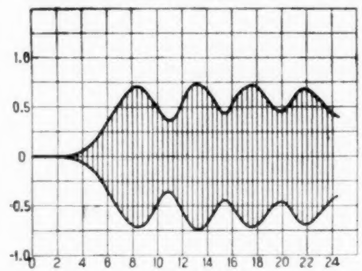


Fig. 42

Envelope of Current in Sixth Section in Response to E.M.F. of Frequency $\frac{1}{2\pi}(\omega_m \pm w/4)$.

High Pass Wave-Filter, Type C_1L_2

Divide ordinates by k and abscissae by ω_c to read current in amperes and time in seconds.

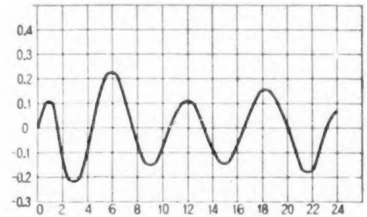


Fig. 43

[Current in First Section in Response to E.M.F. sin $\frac{1}{2}\omega_c t$. Steady-state Amplitude $=0.0415$.

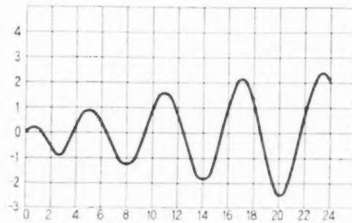


Fig. 44

Current in First Section in Response to E.M.F. sin $\omega_c t$.

Use of Labor-Saving Apparatus in Outside Plant Construction Work

By J. N. KIRK

Introduction

IN the January issue of this Journal was discussed the adaptation of transportation equipment to telephone construction and maintenance work. Closely associated with the operation of such equipment is the problem of utilizing various labor-saving machinery which in many cases has been so designed as to form an integral part of the transportation unit.

It is the purpose of this article to describe some of the more important developments along this line such, for example, as the application of different types of derricks, trailers for various kinds of work, earth boring machines, numerous applications of air compressors and compressed air tools, etc., and in some instances to contrast the latest types of equipment with former manual methods of carrying out similar operations.

Pole Derricks

There are erected in the Bell System each year in the neighborhood of 600,000 new poles. In addition, the maintenance of the existing plant of over 14,000,000 poles involves the moving, removing, resetting and straightening of large numbers of poles annually. This immense task emphasizes the importance of devising means for offsetting, in so far as is practicable, the old manual methods of handling these poles on the job and from point to point in the field as occasion demands.

In 1914 there was developed and put into use a pole derrick of the tripod type which was mounted upon a 5-ton truck from which the derrick received the necessary power for operation. As the use of this derrick, which weighed something over ½ a ton, was extended it became apparent that while the fundamentals of the design and operation were reasonably well adapted to the average construction job, the weight and bulk of the apparatus introduced a very real factor with regard to the available truck capacity. The derrick members, being large and heavy, were difficult for the men to handle and there was not in all cases the desired amount of flexibility to meet the varied and often difficult requirements. This derrick, however, clearly demonstrated the inestimable value of apparatus cap-

able of doing in a few minutes the work ordinarily requiring a large gang of men, many times as long to complete.

An active period of development and experimental field work soon followed the advent of this labor-saving device which has resulted in making available a light type of high grade steel tube derrick.

Figs. 1 and 2 show a pole derrick of the latest type mounted on a $2\frac{1}{2}$ ton truck. Fig. 1 illustrates the method of erecting a pole where the truck can be maneuvered into a position in close proximity to the proposed location of the pole. Fig. 2, on the other hand, shows the possibility of handling a pole at a considerable distance from the location of the truck, which for any reason may be more practicable or desirable.

These illustrations show the derrick in each of the two possible operating positions; in the first instance supported entirely upon the

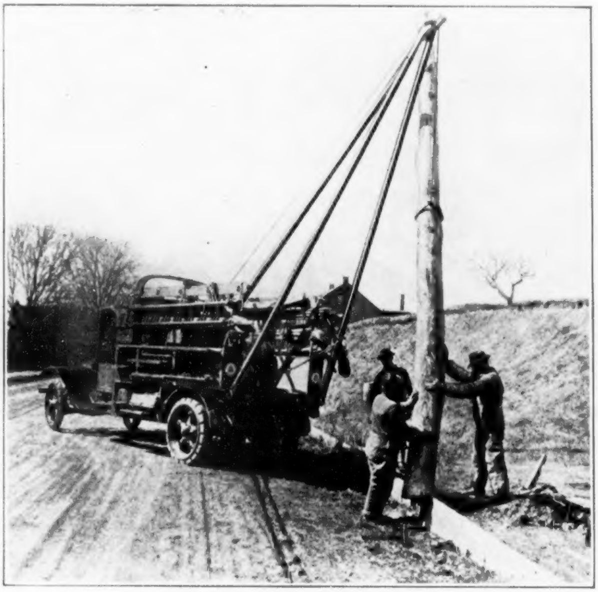


Fig. 1—Erecting Pole, all Derrick Members Mounted on Truck

truck, and in the second, supported from the ground by one of the three pipe members. The derricks of this type are constructed of high grade steel tubing having a strength at the yield point of approximately 70,000 pounds per square inch.

In order that country-wide conditions may be satisfactorily met, the present type of derrick has been made available in two general types which are known as the "middle" and "corner" types for use, as the names imply, from the rear middle or corner of the truck. Each of these types are further available in light and heavy weights, depending upon the lengths and the kinds of poles, cedar or chestnut or other kinds of similar weights, that are generally used in any particular part of the country.

As contrasted with the early type of derrick, the present types weigh from 370 to 520 pounds, depending upon the size used,

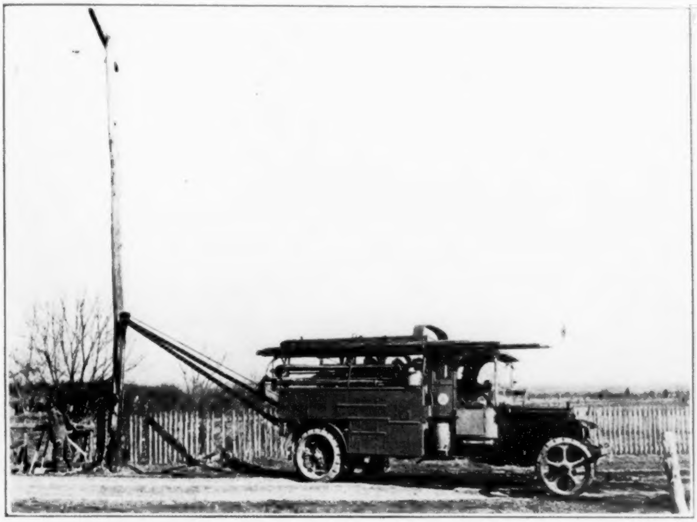


Fig. 2—Erecting Pole at Distance from Truck, One Derrick Member on Ground

and are capable of readily and safely handling any load within the limits of the winch rope capacity, which leaves a satisfactory margin when doing practically any work for which the derrick has a place in telephone construction. Each of the four classes of derricks above mentioned is designed with a view to making its operation as rapid as is consistent with safety. The chauffeur and one man can remove the derrick members from the carrying racks provided on the truck, assemble them and erect the derrick ready for work in from three to four minutes. The disassembling of the derrick requires about the same length of time.

Naturally, the greatest economies may be made in the application of this apparatus where the poles to be handled constitute a consecutive line, the holes for which have been dug in advance. However, because of the short time required for assembling and taking down the derrick, it is generally economical to use it for placing only one or two poles at a location. As indicative of the possibilities with regard to rapidity of operation, it may be of interest to note that in erecting a number of 30 to 35 foot poles under average conditions in a line for which the holes had previously been prepared, a gang of three men have averaged approximately two minutes per pole erected but not tamped.

The use of the derrick has thus far been described as applied to the economical erection of poles. There are, as a matter of fact, many other important uses for which the winch-operated, derrick equipped truck is well adapted, a few of which are enumerated below.

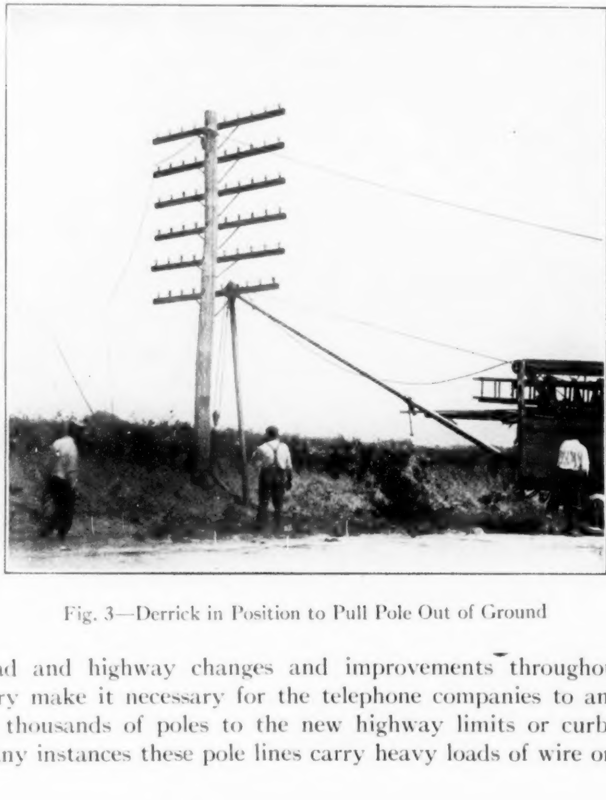


Fig. 3—Derrick in Position to Pull Pole Out of Ground

Road and highway changes and improvements throughout the country make it necessary for the telephone companies to annually move thousands of poles to the new highway limits or curb lines. In many instances these pole lines carry heavy loads of wire or cable

or both. With the pole derrick many of these moves can readily be accomplished without in any way disturbing the wire or cable loads. The derrick pulls the pole out of the ground and with the aid of the truck, the pole with its load intact is moved to the new location where it is lowered into the hole prepared without even untying a wire or loosening a cable clamp. It will also be readily appreciated that the rehandling of cable and particularly the untying of open wires is not only an expensive operation in point of first cost, but that each such operation is distinctly detrimental to the plant, shortening its life and greatly increasing maintenance expenses. It will be

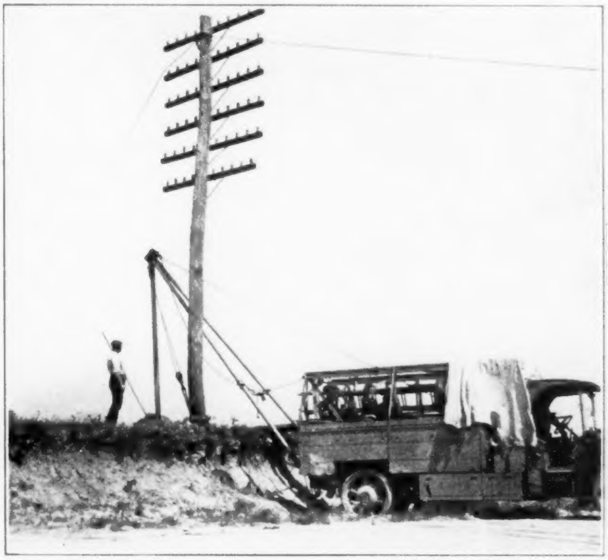


Fig. 4—In Position to Shift Pole to New Location. Pole Has Been Moved Over Bank with Wires Intact

seen, therefore, that the use of the derrick where practicable in connection with the moving of existing lines will largely eliminate the undesirable and costly procedure which is involved in the manual handling of the poles.

As an example of one of the many uses to which the pole derrick can very satisfactorily be put, Figs. 3 and 4 illustrate the initial and final steps in moving back a pole in a 6-arm lead of wires and lifting

it up an embankment to its new location in connection with highway widening. This particular line is about 60 miles long and the distance the poles were moved varied between 6 and 125 feet. It is reported that the move of this entire lead which averaged about 4 arms was completed without untying a single wire, without cutting any slack and with practically no trouble on the circuits. It is needless to say that the saving involved by being able to move this line rather than rebuild at the new location was an item of considerable importance.

The above illustration shows the derrick in position to pull a pole out of the ground, the top of the pole being temporarily side-guyed.

In Fig. 4 the pole is shown after having been pulled out of the ground and placed on top of the embankment. The derrick is ready to shift and slide the pole back to the new hole. Two men and the chauffeur pulled and completed the moving of this pole with its load of six arms of wires in twenty-five minutes.

Fig. 5—Derrick Operating Under Difficult Conditions

As a further example of the usefulness of the derrick in pole work, Fig. 5 shows a job where the pole derrick was operated under rather unusual conditions to erect a pole at the side of the road where the pole hole was dug under water and the pole erected in barrels. It

would be difficult to pike a pole into such a hole because there is nothing against which to rest the butt while raising it.

Another important function of the derrick is that in connection with the resetting of poles or the removal of abandoned poles when it is necessary to remove the butts. The slow and laborious process of pulling the pole out of the ground with a jack or other equipment is practically eliminated as the derrick, properly handled, is capable of doing the greater part of this work in much less time, more economically and with greater safety to the men.

In addition, it might be pointed out that the derrick equipped truck is also becoming more and more indispensable in connection with the handling or moving of any heavy loads in the storage yards, in unloading or in moving stock supplies of poles under adverse conditions and many other uses.

Fig. 6—Erecting Pole by Manual Methods. Contrast with Previous Operations

In contrast with the mechanical methods of erecting and handling poles as previously shown, Fig. 6 shows the old manual method of erecting a large pole. Not only is the number of men required large,

but the observance of most rigid precautions does not entirely preclude the possibility of hazard to the men when handling the heavier poles. Further, the pole locations are not always such that a considerable number of men with pikes can properly distribute themselves about the pole so as to complete the raising and lowering operations in a reasonably safe and efficient manner.

Earth Boring Machines

One of the slowest and most difficult physical tasks connected with outside construction work is that of digging pole holes. It is estimated that upwards of 1,000,000 holes must be dug annually to accommodate the poles erected in new locations, and those replaced, moved and reset in the Bell System. Under soil conditions reasonably free from obstructions a man can generally average about three holes per day with perhaps five to six as a maximum under ideal soil conditions, while in more difficult digging one or possibly two holes may represent a good average day's work. It probably requires somewhere in the neighborhood of 3,500,000 man-hours per year simply to dig pole holes.

For a number of years the availability of a practical pole hole digger has been the objective of telephone linemen. Development work has progressed rapidly during recent years and the high point of perfection which has been reached in automobile truck design and performance has greatly simplified the adaptation and increased the practicability of the boring apparatus. It is of interest to note in this connection that the solution of the problem comes at a time when there is a pronounced shortage of common labor.

The construction in 1914 of that portion of the transcontinental line extending across Nevada, marks the first really economical application of a machine to bore pole holes. In about 1917 the need for labor relief led to renewed activity in connection with adapting the fundamental principles of the original boring apparatus to machines sufficiently flexible to meet the general and rather exacting requirements of telephone work.

Fig. 7 shows one of the latest developments in earth boring machines, which is cleancut and rugged. This machine is mounted upon a 4-wheel drive truck and is otherwise specially equipped which enables it to reach practically any location where it is necessary to bore holes for the erection of poles. As a matter of fact it has been demonstrated that these machines are able to reach approximately 95% of the pole locations. Further, the machine being equipped

with a pole raising derrick makes possible the digging of the hole and the erecting of the pole with but one setting of the truck.

With the boring machine from 30 to 80 poles per day can be set in their holes by a force of three men. This, of course, does not include straightening the poles and backfilling the holes. To do this amount of work with manual labor only would ordinarily require from 15 to 50 men. It is of particular interest to note that the more difficult the digging, exclusive of rock, of course, the greater the saving by using

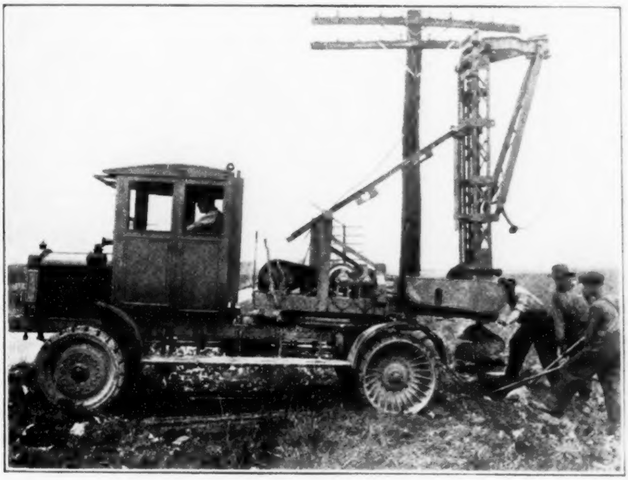


Fig. 7—Boring Hole for "H" Fixture

the machine. It might be mentioned that one of the most important features of the boring machine is its ability to bore holes through frost thus enabling a more uniform apportionment of pole work over the entire year. This feature is also of particular value in connection with the restoration of service subsequent to sleet storm breaks in winter at which time hand digging is in many cases a practical impossibility.

Fig. 8 illustrates the ability of this 4-wheel drive outfit to negotiate difficult ground conditions. In this instance one rear wheel has dropped into a hole while traveling over a plowed field covered with snow. It required only a few minutes to lift the wheel by moving the turn table so that the auger was just behind the buried wheel, then raising that corner of the truck by forcing down the auger with

power from the engine, sliding a skid board under the wheel thus raised, lowering the wheel to this board and driving away.

Fig. 8—Machine Extricating Itself from Hole

Cable Reel Trailers

To meet the need for a device suitable for trailing a single reel of cable and also for use as a reel "set-up" preparatory to a "pull" of either underground or aerial cable, a type of cable reel trailer has been developed as illustrated in Figs. 9 and 10.

A number of trailers of this type have been in service for some length of time and their use has brought out many advantages, some of the more important of which are:

> A reel of cable can be loaded on and unloaded from the trailer in less time and with less effort than when a reel is carried in the body of the truck. In this connection, it might be pointed out that an important safety feature is involved in that the hazards to the men in loading and unloading heavy reels of cable by the old method are practically eliminated. Of course, even where reels of cable are carried in the truck the use of the winch and spindle as previously discussed in the January issue eliminates the hazard that was present in the old method of loading and unloading, involving the use of skids.

Fig. 9—Truck Being Used to Load Reel of Cable on Trailer

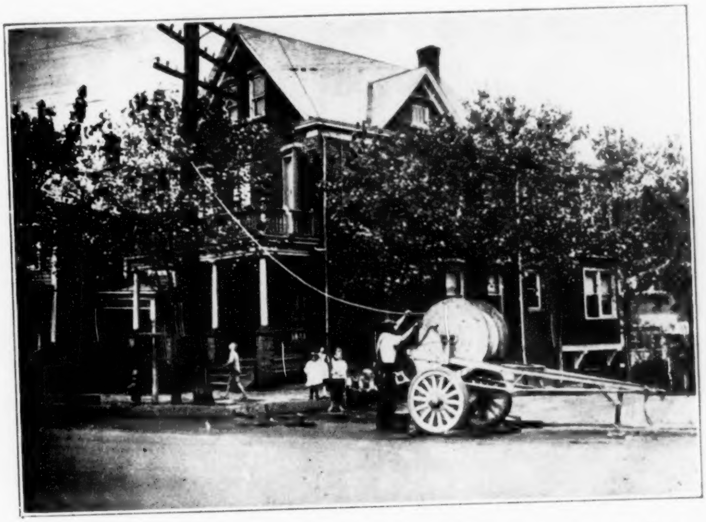


Fig. 10—Cable Being Pulled into Rings from Reel "Set-up" on Trailer

Fewer men are required for loading, unloading and "setting-up." For example, two men with a chauffeur and truck (not necessarily equipped with a winch) can satisfactorily handle a 3-ton reel of cable with the trailer, where ground conditions are such that they can maneuver the reel on the ground.

Where a single reel of cable is to be used for one "pull" or for a number of short "pulls," the trailer is used to haul the reel to the job and to "set up" the reel for each "pull." The reel may be trailed, in addition to carrying materials, tools, etc., in the body of the truck, thus making it unnecessary to unload or disarrange the equipment regularly carried on the truck.

When delivering a number of reels, one reel may be trailed in addition to carrying one or more on the body of the truck, thus materially increasing the hauling capacity of the truck, with a proportionate reduction in delivery costs.

As the photographs indicate, these trailers are equipped with springs and rubber tires which afford material protection to the cable while in transit.

Pole Trailers

For the transportation of poles under ordinary conditions, the use of a two-wheel trailer with the poles balanced on the trailer and towed behind the truck is ordinarily the most satisfactory method. Fig. 11 shows such a trailer loaded and ready for action. This method has the advantage that the trailer loaded with poles can be readily detached from the truck and left at any desired location, thus releasing the truck for other work. Also, in case of the load being stuck on a hill or in the mud, the trailer can be readily detached while the truck runs forward and from the top of the hill or from firm ground, pulls the trailer load of poles by means of the winch line.

Limiting the weight to conform with requirements of state laws materially limits the size of the load in hauling chestnut and creosoted pine poles. However, in the case of cedar poles, the bulk of the load rather than its weight is ordinarily the limiting factor.

To meet these different conditions, three sizes of pole trailers have been designed, a heavy duty trailer rated at about 8 tons with ample overload capacity, a medium duty trailer rated at 5 tons, and a light duty trailer of $2\frac{1}{2}$ ton capacity for use in districts where it is desirable to maintain a standard tread between the wheels rather than to use the narrow tread dinkeys for the lighter pole loads.

Fig. 11—Balanced Load of Chestnut Poles on Trailer

Block Gang Trailer

Fig. 12 illustrates a type of trailer which has been developed recently for the use of gangs doing interior block construction work. In a

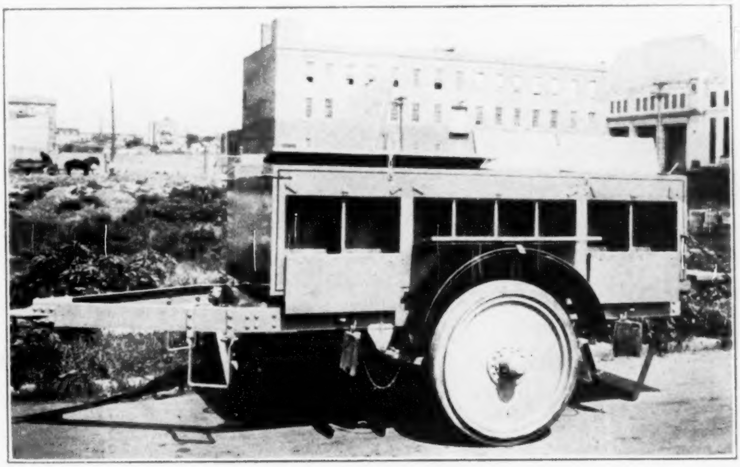


Fig. 12—Trailer Equipped with Special Body for Interior Block Construction Work

case of this kind, the gang is ordinarily located on a job from one-half day to three or four days, and since the power equipment on a truck would be of no value in connection with placing a cable on the rear walls of buildings, for instance, it is more economical to serve this gang by means of a trailer.

This light type of trailer contains sufficient space for carrying all the necessary miscellaneous tools and materials required in connection with block work and the compartments into which it is divided are such that these articles can be arranged in an orderly and readily accessible manner, thus making for increased efficiency in executing the work.

Concrete Mixers

In connection with the construction of underground conduit and particularly in the work of building concrete manholes, which are now being employed to a rather large extent, it is essential that concrete mixers be available which will be especially adapted to telephone work. Some of the requirements of this service are that the outfit be of light weight, compact, embody maximum portability, and be reliable in operation. The failure of a mixer on a telephone job may seriously handicap the operations of a large gang of men.

Fig. 13 shows a commercial type of mixer which has been modified

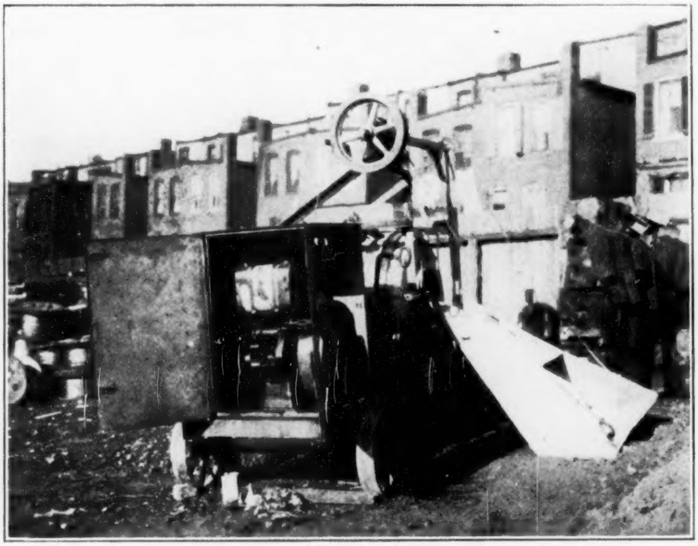


Fig. 13—Concrete Mixer Adapted to Meet Telephone Construction Requirement

in several respects to meet the particular requirements of telephone construction work.

Units of this type which are now in service are operating very satisfactorily, both from the viewpoint of reliability and adaptability to the work. This outfit will mix as much concerte as ten men and will do it much better.

Fig. 14—Pouring Concrete Manhole. Note 4-way Chute for Distribution

Fig. 14 shows one of the batch mixers in service pouring a concrete manhole, the concrete being uniformly distributed to all sides of the structure by means of a four-way chute. In connection with the broadening use of concrete manholes it might be mentioned that the availability of improved compressed air tools has greatly simplified and cheapened the making of any changes that may be required subsequent to the initial construction of the manholes.

In order to provide a concrete mixer unit having maximum portability and having proper capacity and operating features for telephone work, we have cooperated with the manufacturer in the development of such a unit which is shown in Fig. 15. This consists of a batch mixer permanently mounted upon a Ford 1-ton truck chassis and operated through a suitable power take-off from the Ford engine. This unit loads from the ground by means of a power loader and distributes the concrete from the opposite side of the drum through a long swinging adjustable chute (not shown). A small trailer if desired

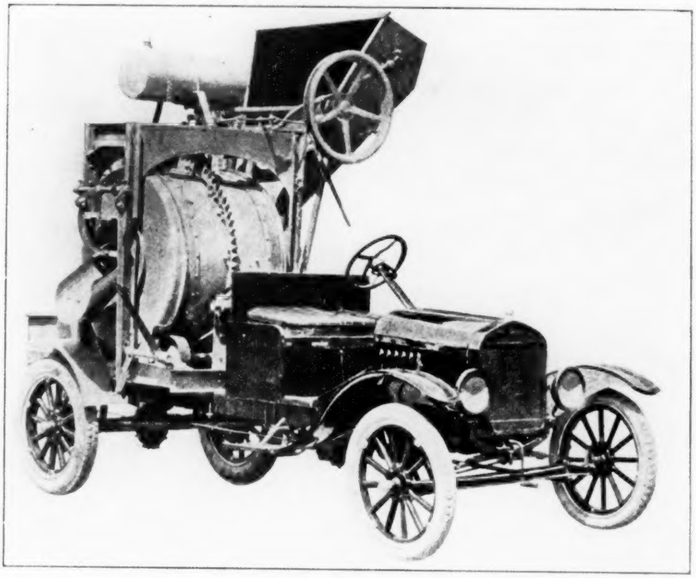


Fig. 15—Concrete Mixer on Ford One-Ton Truck. Maximum Portability for Small Jobs

Fig. 16—Light Weight Trenching Machine.

can be used behind the Ford truck to transport the supplies and tools necessary in connection with isolated jobs.

Trenching Machines

Where it is necessary to do a considerable amount of trenching for underground conduit in outlying districts, it is sometimes possible to utilize a trenching machine with marked economy. In fact under normal conditions a machine of this kind will dig trench about as fast as a gang of 50 men.

The machine shown in Fig. 16 is a recent development which has advantages over the older type units in that the size and weight are such as to admit of its being transported from point to point on a heavy truck or trailer.

Pumps

In handling the water from excavations and also from manholes where splicers are working, the diaphragm pump illustrated in Fig. 17 is giving a good account of itself, particularly because of certain features incorporated in the design which especially adapt it to telephone conditions.

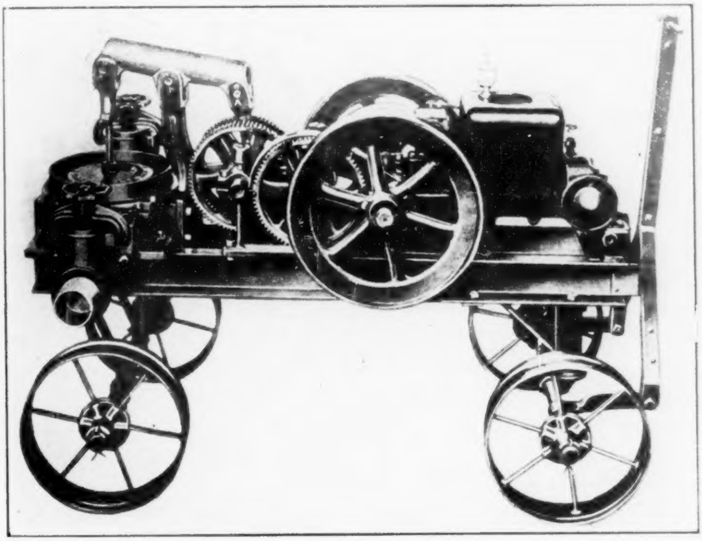


Fig. 17—Enclosed Discharge Diaphragm Pump. Capacity One Barrel per Minute

This little unit will pump water at the rate of over one barrel per minute and discharge it through a hose away from the job to any location desired. It will operate all day with practically no attention, upon a gallon or two of gasoline. When pumping under ordinary conditions it will handle water faster than 12 men with hand pumps.

One very desirable feature of the diaphragm pump is that it is self-priming. For instance, if splicers are working in a manhole the pump can be started and the initial volume of water removed, then as seepage water enters the manhole it will be immediately taken up and discharged without any attention from the splicers or helpers.

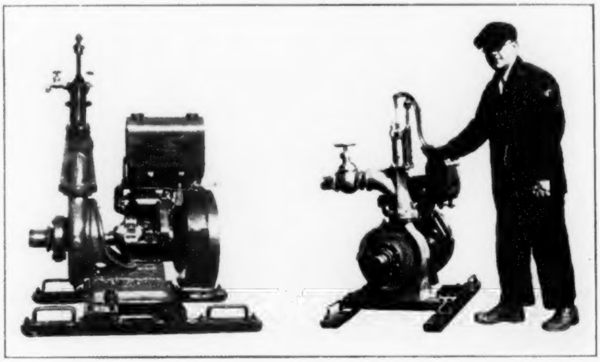


Fig. 18—Light Weight Centrifugal Pump. Capacity Seven Barrels per Minute

For handling larger volumes of water there has just been developed, as the result of careful study and cooperation with the manufacturer, a new type of centrifugal pump shown in Fig. 18. This unit consists of an air cooled engine similar to that used in the concrete mixers. On the end of the engine shaft is mounted the centrifugal pump impeller. The pump casting also forms a base for the engine.

As an indication of the capacity of this pump it might be of interest to note that it would not be possible to get enough men with hand pumps around a manhole to handle water as fast as this unit. It will pump seven barrels of water per minute and mounted on skids as shown it weighs only about 500 pounds.

In the case of trucks which do a considerable amount of underground cable placing in districts where water must be removed from manholes in advance of the cable placing operation, centrifugal

pump equipment mounted on the truck is desirable. As soon as the gang arrives at a wet manhole, the pump if promptly applied will remove the water in the few minutes during which preparations are being made for placing the cable, so that ordinarily the gang is not delayed in the least by the water.

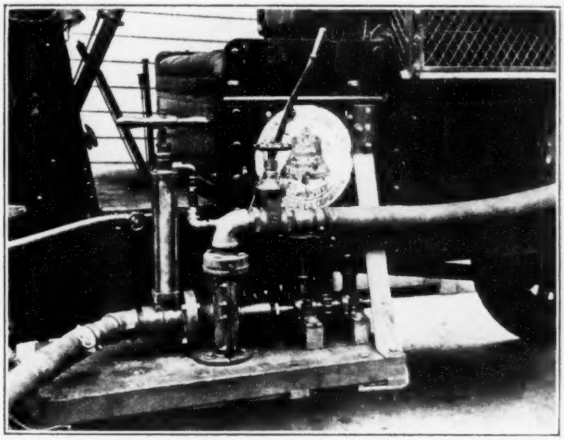


Fig. 19—Centrifugal Pump Mounted on Underground Cable Placing Truck

There are several points in favor of locating the pump on the running board as shown in Fig. 19 rather than in the body at the rear of the cab as has been the usual practice in the past. With the running board installation the water is not carried up into the truck body where it has a tendency to get into the tool and material boxes and equipment and also to cause deterioration of the body. In addition space is economized and the pump is located considerably lower than would otherwise be the case, thus resulting in a reduction of the suction lift for the water between its level in the manhole and the pump impeller.

Air Compressors and Compressed Air Tools

Of the many applications for mechanical equipment to offset the scarcity and high cost of certain types of labor such as for excavating, etc., the use of air compressors and compressed air tools is of prime importance in the outside plant construction work. Through special adaptations to meet each peculiar condition, this class of labor saving equipment has been made available for use on such jobs as the opening

of all kinds of street pavements preparatory to laying underground conduit, cutting frozen ground, loosening the earth in excavating instead of using picks, drilling rock preparatory to blasting for underground conduit or for pole holes, tamping back filled earth, cutting iron pipe covering from cable, etc.

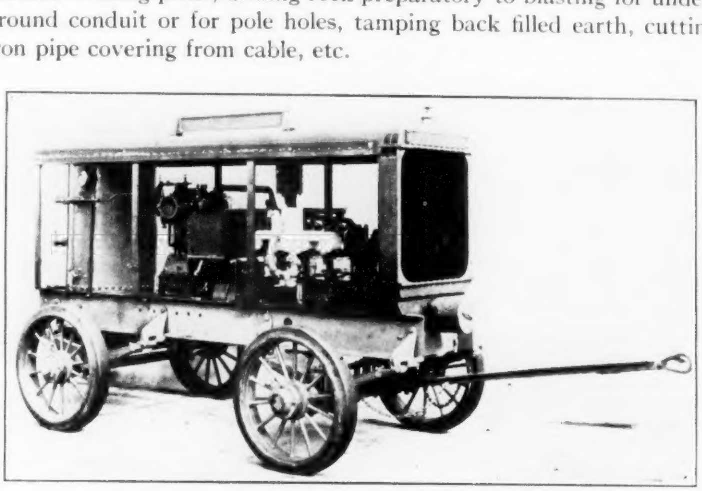


Fig. 20—Air Compressor Mounted on Trailer for Maximum Portability

Fig. 20 shows a new type of portable gasoline engine driven compressor unit which is being satisfactorily used for the larger jobs of opening street pavements, for rock drilling, etc.

Fig. 21—Removing Granite Blocks and Breaking Concrete Base

Where trenching work involves cutting through paved streets one compressor unit with three men will ordinarily accomplish as much in a given period of time as 27 men employing former methods.

In Fig. 21 two operators are shown opening pavement which consists of granite blocks set in cement, on a concrete base. One man goes ahead and wedges the blocks loose, while the man following breaks the concrete base. Some pavements of this type are very difficult to open. When the cement filling is of good quality the granite blocks often break before the cement loosens.

Fig. 22—Air Gun Cutting Asphalt

Fig. 22 shows an operator cutting asphalt pavement. With the wedge-shaped blade cutting at intervals as shown, small cracks are opened between the holes so that when cross cuts are made square blocks of asphalt can be readily lifted out.

The above illustrations contrast rather strikingly with the old methods of cutting pavements by means of sledges and bars as shown in Figs. 23 and 24.

In the use of the old manual method of cutting with sledges and bars there is always present a certain degree of hazard to the men. There is the possibility of the striker missing the steel and striking the holder's wrist, also the danger to the men's eyes from flying steel chips. These safety points, of course, are outside the labor saving considerations.

Fig. 23—Manual Method of Breaking Concrete. Contrast with Fig. 21

Fig. 24—Manual Method of Cutting Asphalt. Contrast with Fig. 22

While the labor saving is large in connection with opening street pavements, it is even greater in the work of drilling rock for blasting, where two men and a compressor can ordinarily do as much work in a given length of time as 35 to 40 men using hand methods.

In Fig. 25 is shown another interesting and efficient application of compressed air tools. Compressed air spades are being used to an increasing extent for loosening hard earth instead of doing this work by the usual pick method. A tool of this kind requires very little air and while this particular application is rather new, it is felt that further study may result in an appreciable saving over hand pick methods.

Fig. 25—Pneumatic Spade Replacing Hand Pick Method of Loosening Hard Soil

Compressed air can also be used to advantage in tamping back filled earth. Under certain conditions, however, it now seems that a suitable mechanically operated tamper will probably show greater economy on all except jobs in congested areas where the underground pipe interference is serious or where the trench or opening extends in a diagonal direction, thus often precluding the use of a rigid mechanical device.

The utilization of the portable air compressor is a comparatively recent development undertaken by the telephone companies in co-operation with one of the large air compressor manufacturers.

While the large capacity units have reached the stage where they give satisfactory operation, there is a field in the telephone business for a much more compact, lighter weight unit of lower capacity and cost, for such work as the opening of trench for subsidiaries, cutting frost, drilling rock for pole hole blasting, etc. With this in mind there has recently been developed in cooperation with an air compressor manufacturer, a type of compressor which is suitable for operating either one jack hammer for rock drilling or one tool for street opening with a corresponding capacity for other types of compressed air work. It is expected that the weight of this unit can through further study be reduced to such an extend that it will be practicable to mount it upon a Ford one-ton truck and still leave sufficient carrying capacity to handle the necessary guns, steels and hose for operating. Where there will be practically constant use for this lighter unit it may be desirable to mount it permanently upon the truck, while, in cases where the use will be intermittent, a very economical and convenient mounting can be made upon one of the Army type trailers.

Conclusion

In this article an endeavor has been made to cover in a very brief way some of the more important items of mechanical application which have a place in telephone construction work. The adaptation of mechanically operated tools and other devices to assist in the necessary manual operations will undoubtedly continue to occupy an important place in the work. Further study and development should result in many improvements in the present-day way of doing things which will make not only for marked economies of operation, but for greatly increased features of safety to the men engaged in constructing and maintaining the telephone plant.

A Method of Graphical Analysis

By HELENE C. BATEMAN

Introduction

IN connection with many telephone problems of an economic character, it is necessary to develop methods for making estimates and forecasts of the effects of changes in conditions. When the changes in conditions are such that direct experimentation is impracticable the development of logical methods and bases for estimates involves analyses of past experience in specific situations and, in so far as is feasible, the generalization of such experience. It is the purpose of this paper to describe briefly a graphical method by which complex economic data may be generalized for use in forecasting probable future conditions.

In some problems, it is necessary to determine the effects of changes in a specific situation, the results being applicable particularly to the given situation, and only very generally to other situations. The effect of a change in population upon station growth in a given exchange is an example of such a problem. In other problems, it is practicable to generalize experience so that the results of analyses may be applied, under proper conditions and limitations, to various specific situations. Moreover, it is often necessary to apply a general conclusion to a specific situation because no specific experience is available. An example of this type of analysis is the generalization of results of various rate treatments in different exchanges. In meeting this type of problem graphical methods are utilized to compare experience of a similar nature in various situations. The factors which may be indices of differences in conditions among various situations are studied to determine their relation to the differences encountered. Finally an attempt is made to derive quantitative relationships from the experience analyzed.

The assumption made in utilizing such methods is that the experience in different situations, from which generalizations are to be made, is *essentially similar* in certain respects, and that the variation in the quantitative unit to be estimated is due to varying conditions, as between the different situations, which may be measured in part by quantitative factors. There are, of course, certain types of problems where essential similarity between different situations does not exist or where it is difficult, if not impossible, to isolate quantitative factors sufficiently reliable to form a basis for estimates. On the

other hand, there are many problems to which these methods may properly be applied and in which it is practically impossible to derive a reliable and satisfactory basis for making estimates without some such methods of analysis. Certain economic problems, in particular, because of the impracticability of experimentation and because the complex reactions of a group of individuals are involved, are not adapted to solution by the statistical methods which have proved useful in biometric sciences, but may be dealt with by graphical methods. This has been found particularly true in problems involving local telephone message use, and throughout the following discussion, illustrations are drawn from analyses of this type.

Data

Since the ultimate aim of a graphical analysis of this type is to provide a basis for making estimates, the first step is to determine the estimates which will be required and the type of cases and conditions under which they will be used. In this way the aim and scope of the analysis is clearly defined. The unknown factor (the dependent variable) is to be estimated from certain known factors (independent variables). Various factors, quantitative and qualitative, which might logically appear to be indices of conditions controlling the dependent variable are, therefore, considered.[1] Only factors as to which data are available at the time and place where estimates are to be made are useful as independent variables. It is usually advisable to test a suggested factor by means of any data, even in small amounts, which may be available before a complete body of data is collected. Such preliminary investigations are useful in indicating the scope and detail in which data should be secured. In general the data should:

1. Be adequately representative of the type of cases for which estimates must be made,
2. Be adequate from a sampling standpoint for each situation,
3. Be as nearly homogeneous as practicable, i.e., cases having any outstanding peculiarities should be excluded,[2]
4. Include what appear to be the important factors or indices for each case.

[1] It should be noted that such relationships need not be those of cause and effect. If two factors vary together (as do, for instance, different effects of a common cause) the values of the one which are hard to determine can be estimated from the more easily measured values of the other.

[2] For instance, if estimates are to be made for small exchanges, it would not be advisable to include data from large exchanges in the analysis.

Preliminary Analysis

After the data have been collected and summarized in accordance with the general plan of the study, the graphical phase of the analysis begins with trial setups in which the dependent variable is plotted against each of the independent variables in turn. Such charts are intended only to give a general idea of the types of relationships and to determine which of the factors tested are most closely related to the dependent variable. Factors which do not vary with the dependent variable are not necessarily to be discarded permanently since the effect of one factor may obscure that of another. It is not to be expected that the data plotted on any of these charts will fall along smooth curves. They will probably be widely scattered but in the case of the more important factors a general trend is usually evident.

On the next series of trial charts, several of the more important factors are considered simultaneously. If a qualitative factor is under consideration, separate charts are plotted for the different classes. If these charts are essentially similar, the qualitative factor may be disregarded for the time being and the data considered as a whole. If, however, the qualitative factor appears to influence the relationships in a logical manner the data must be sub-divided and a number of practically independent studies carried on. In fact, the analysis of the effect of a qualitative factor is intended to determine whether or not the data forms an essentially homogeneous whole. If there is a discontinous variable, it is often convenient to hold it constant, i.e., a separate chart may be plotted for each value or group of values of this factor. The factor, which from the preliminary charts, seems most important is usually plotted against the dependent variable. One or two other factors are coded. The codes may be either in colors or symbols or both. The color codes are usually the more easily distinguished and are, therefore, the better for working charts. For final charts, however, color codes are not usually practicable because of the difficulties of reproduction. Both colors and symbols may be used when two coded factors are to be tested simultaneously.

In these preliminary sets of charts, it is well to test as many different factors and combinations of factors as appear logically to vary with the dependent variable. It is usually best, however, to consider not more than three or four independent variables at a time, one plotted against the dependent variable with one or possibly two coded and one held constant on each chart. An attempt to hold constant a greater number will often sub-classify the number of data points so far as to obscure the real trends. Furthermore, the com-

plexity of charts increases rapidly with the inclusion of more variables and makes the analysis and estimating complex and cumbersome.

Fig. 1 is a typical preliminary trial chart from a study of average telephone message use under message rate service. Each data point represents one class of service in a particular exchange. The independent factors taken into account are:

1. Major Service Classifications[3]—held constant since this chart is for one class only.
2. Rank of Service[4]—plotted.
3. Message Allowance—coded.

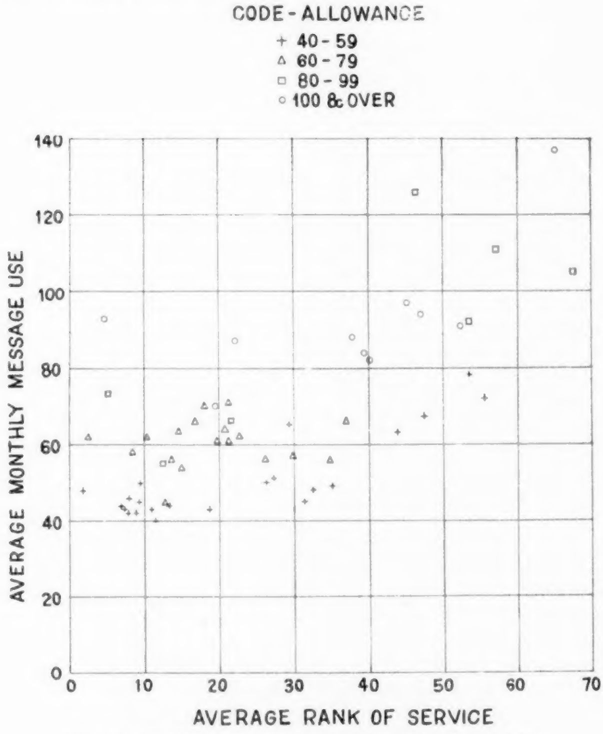


Fig. 1—Preliminary Chart Involving Three Variables

The trend of the relationships between message allowance, rank of service and average message use is fairly well defined on this chart.

[3] Business Main Station, Residence Main Station, P. B. X., etc.

[4] A statistical index indicating the relative ranking of subscribers in accordance with their demands for service.

When several different sets of charts such as are described above have been scrutinized, definite trends will usually be fairly clearly established. It will often be found that while these trends are well defined, nevertheless a number of points may scatter widely. Such points are studied carefully. If, after the original data are checked, the points are found to be correctly plotted, each case is investigated in detail to account for the observed divergence. Sometimes it will be found due to a factor which has not been taken into account, the inclusion of which will often improve the results of the study as a whole. On the other hand, peculiar local conditions or history may give rise to such divergence. These cases are not really a part of the similar group under consideration. If they are sufficient in number and similar with respect to each other they may be studied independently. If not, they are either excluded entirely or given slight weight in the general study. Because of wide differences in problems and material, it is not practicable to describe in detail the process of analyzing such preliminary charts in arriving at decisions as to data and process.

Curve Drawing

The next step is the construction of curves through these data which will truly represent the relationships involved. This can be facilitated by plotting the average values of the dependent variable for all cases having the same values (within certain limits) for all the independent variables.

> On Fig. 2 the data points are the same as those plotted on Fig. 1. The closed symbols which have been added are average points representing the data points of the same symbol. The abscissa of each average point is the mid-point of an interval of rank of service (0–20, 10–30, etc.) and the ordinate is the average of the message uses of all the points falling within that interval.

The average most often used on such charts is the median not only because it is most easily located but because it is usually the most representative, giving little weight to extreme cases. Whatever average is used, it is well to make it a moving average, i.e., covering overlapping intervals such as 20–30, 25–35, 30–40, etc., rather than 20–30, 30–40, 40–50, etc.

These averages serve as a guide for drawing the preliminary curves through the data but the actual data points are considered at the

same time. In constructing the curves, the significance to be attached to any data point depends chiefly on:

1. The number of individual cases on which it is based.
2. The probable degree of accuracy of the data.

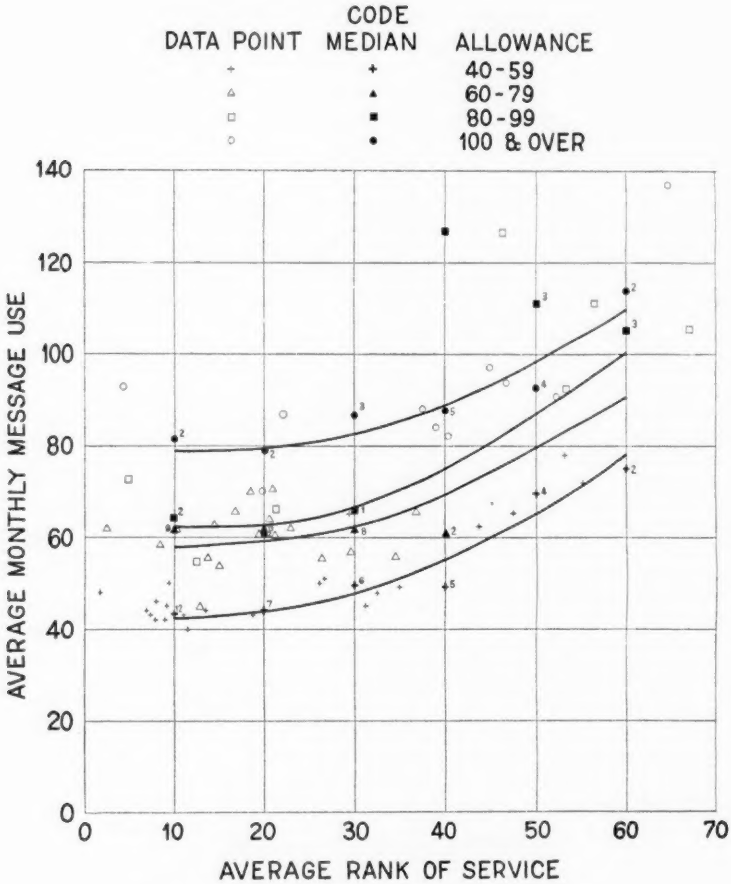


Fig. 2—Preliminary Chart With Averages and Curves

Since these characteristics are considered simultaneously it is usually advisable to depend on judgment using the averages as a general guide, rather than to rely on any formal mathematical system. The first set

of curves is drawn to fit the data as closely as practicable and still be *logical* and *consistent.*

On Fig. 2, the number of data points on which each average is based has been noted as an aid to judgment and a set of rough preliminary curves has been drawn. These, of course, are not necessarily the most accurate curves which could be constructed from these data. A method of progressing to final curves is described below.

Curve Smoothing

The first set of curves constructed from the data may not be an entirely consistent and reasonable family. The relation between different curves on the same chart or between different charts indicates the influence of factors other than the one plotted and must, therefore, be made consistent and logical. The process of transforming the preliminary curves into the final normals is known as smoothing.

The original curves are first studied for reasonableness. Their general shape (whether straight line, convex or concave, having maximum or minimum points, being asymptotic to a certain line, etc.) is, in so far as practicable, determined on logical grounds. If a large majority of the curves, or the curves based on the greatest amount of data, have a certain clearly defined trend, the remainder of the curves are made to conform to this trend, if it is reasonable, at the same time keeping as closely in line with the data as possible.

Each chart will usually have one independent variable plotted against the dependent variable and another independent variable coded. Each curve, therefore, indicates the relationship between the dependent variable and one independent variable for a certain constant value or range of values of a second independent variable. If the relative positions of the curves of a family on one chart are adjusted, the relationship of the coded variable to the other two is altered. The effect of this alteration may be seen by plotting the coded variable against the dependent variable and coding the one which previously was plotted, all values being read from the preliminary curves. This is sometimes called cross-sectioning. The families of curves formed by cross-sectioning are then smoothed until they are reasonable and consistent. In doing this, the original curves are automatically departed from, and when the original curves are replotted from the cross-sections, it may be found that the resulting family of curves is not smooth, consistent or reasonable.

The smoothing process must, therefore, be repeated back and forth a number of times until both sets of curves appear to be smooth, reasonable and consistent families. During this process, it is important that the various families of curves be tested against the original data. If this is not done, it may happen that a series of small changes will accumulate in such a way as to bring portions of the curves outside the limits of the original data. Furthermore, the factor or factors held constant on each chart must not be lost sight of. These factors should be plotted against the dependent variable holding constant

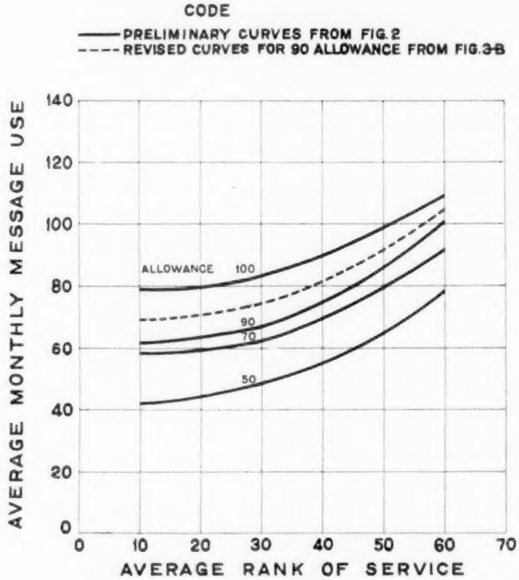


Fig. 3A

all other factors (values being read from the smoothed curves) to see that these relationships also are being made reasonable, consistent and smooth.

The process of smoothing described above is a long and laborious one involving at every step the making of special decisions based upon knowledge of the data and the logic of the situation with regard to the particular problem. Various methods of facilitating the work have, however, been devised some of which are described below. Figs. 3A and 3B illustrate the advantage of having both sets of curves on the same chart with the same scale for the dependent variable so

that when the smoothing process is applied to one family, the effects on the other may be more readily ascertained.

The curves on Fig. 3A are the rough curves which were drawn through the data on Fig. 2. Fig. 3B shows the cross sections of these curves, message allowance being plotted and rank of service coded. It is evident that neither set of curves is a consistent family. Most of the curves of Fig. 3B are irregular instead of being smooth. They might be smoothed considerably either by lowering the points corresponding to a 70 message allowance or by raising those for an al-

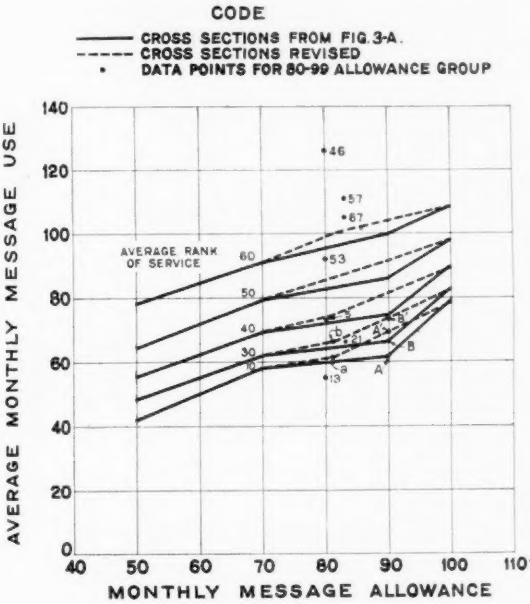


Fig. 3B

lowance of 90 messages. A study of Fig. 3B shows that the curves at 90 allowance are further apart than at any other point. This might be used as an argument either for raising the lower points or lowering the higher. In scrutinizing the data, however, it is found that of the classes of service having from 80 to 99 allowance all have allowances of either 80 or 83. Therefore, the midpoint (90 allowance) is too high to represent the group, or, conversely, the message use plotted is too low for an allowance of 90 messages. In order to have a guide in the amount of shifting necessary, data points for the actual message allowances of the 80–99 group have been plotted

on Fig. 3B, and the values formerly entered at 90 (points A, B, etc.) entered at $81\frac{1}{2}$ (points a, b, etc.). With these points and those at 100 as a guide, new values for 90 allowance have been estimated (points A′, B′, etc.). The shifting of a point up or down on Fig. 3B results in shifting the corresponding points of the other family (Fig. 3A) the same distance in the same direction resulting in the dotted curve. There is much more smoothing necessary to make Figs. 3A and 3B satisfactory and reasonable, but by proceeding in the manner just described, taking into account the appearance of the curves, the logic of the situation and the original data, a smooth and consistent family of curves can finally be evolved.

Another excellent method of smoothing involves the use of a three dimensional figure. Just as a plane surface gives a complete representation of two variables and a partial representation (by coding) of a third, so a three dimensional system can be used to give a complete representation of three variables, and a partial representation of a fourth. It also aids greatly in smoothing simultaneously. A device for three dimensional representations consists of a plane surface marked off with rectilinear coordinates and having at frequent intervals holes into which pegs can be set. The pegs also have coordinate markings. The values of two variables, then determine the point at which the peg is set and the value of the third determines a distance along the peg. The point is marked by a small rubber ring which fits around the peg. The values of a fourth variable may be coded by using rings of different colors. When the device is used for smoothing curves involving only three variables, the data points may be indicated in one color and the smoothed values in another. The data points, remaining constant as the smoothed curves are shifted, form a continuous check on the divergence of the smoothed curves from the data. This is an automatic process of cross-sectioning. When the position of a point is changed, the effects of the change on the various relationships are seen by studying different aspects of the setup. This device gives the best results with discontinuous variables (such as message allowances, rates, etc.) as the pegs can then be set in at regular intervals without resorting and regrouping the data. It is also especially valuable when one of the variables is a complex factor (such as distribution of development among more than two classes of service) which cannot easily be represented by one curve.

Fig. 3C illustrates such a setup with the revised curves of Figs. 3A and 3B. The independent variables, message allowance and rank of service are represented by the rectilinear co-

ordinates of the plane surface. Message use is indicated by the rings on the upright pegs.

In general, the various steps in the analysis leading up to the final[5] or normal[6] relationships require continuous exercise of judgment. The problem is never one of securing simply curves of "best fit" to the data. It is broader, more fundamental and much more involved

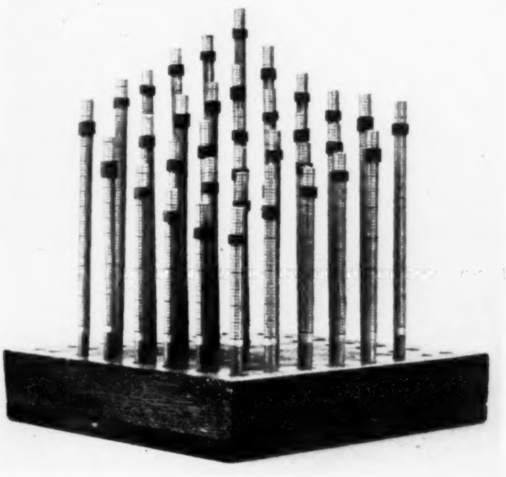


Fig. 3C

than this. It requires a combination of logic with the data that results in normal relationships which fit the data and at the same time are reasonable. It is necessary to consider such questions as the following: Why do the data indicate this relationship? As a generalization, is such a relationship reasonable? What should be the character of this relationship? Should it be a straight line, concave up or concave down? Particular attention is given to the reasonableness of maxima or minima points and to points of inflexion when indicated by the data. It is only by considering such fundamental questions that a sound basis can be established for building up normal rela-

[5] Final in a relative sense. In economic studies of this type involving human reactions and relationships normal relationships are never final in an absolute sense.

[6] The term "normal curve" is used throughout this paper to designate a final curve from which estimates are to be made. A normal distribution curve in this sense may or may not be "normal" in the statistical sense of an evenly balanced bell-shaped curve.

tionships which will be a true generalization of experience. The importance of dealing with economic problems in this way can hardly be over emphasized. It is a recognition of the complexities which are inherent in problems involving human reactions and the dangers of untrue generalization if rigorous and more or less inflexible methods of analysis are utilized.

Final Results

The result of the smoothing process is the development of a consistent series of charts and curves by means of which the value of the dependent variable may be estimated from the values of the various independent variables.

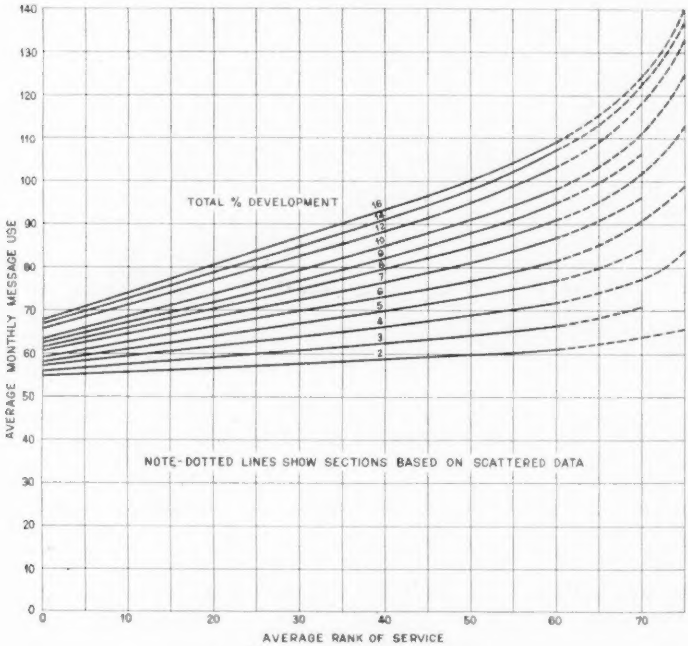


Fig. 4—Final Family of Curves

Fig. 4 is an example of such a chart for estimating average message use. Charts of this type may be used under proper conditions for estimating either an actual value which is unknown (such as average message use under an existing rate schedule) or the value which may be expected to result from some change (such as average message use under a proposed rate schedule).

After deriving a series of final charts estimates are made of the value of the dependent variable for all the cases on which the study was based. Consideration of the differences between the estimated and the actual values is an excellent general criterion of the accuracy of the normals. In general, the positive deviations should be approximately equal to the negative both in number and in the sum of their numerical values. If either positive or negative deviations are decidedly predominant, it is probable that the general level of the normal curves is too low or too high.

When the deviations (without regard to sign) are plotted as a frequency curve, the curve should be fairly smooth. It need not be and usually is not a bell shaped curve, but if there are sudden and decided breaks, it is probable that either certain portions of the data have not been given proper consideration or that the data were not originally essentially homogeneous. The cumulative frequency curve based on the deviations makes possible the easy reading of the median or probable error of estimate. The probable error may be used as a general criterion of the value of future estimates made from these normals and the ratio of the probable error to the median value of the dependent variable forms a basis for comparison of the relative accuracy of different sets of normals.

The deviations (sign being taken into account) when plotted against the various factors included in the study should be fairly evenly scattered and show no trend or relationship. If a consistently occurring variation is discovered between the deviations and any of the independent variables it indicates that the relationship of that variable to the dependent variable has not been properly taken into account in deriving the normals. If this variation appears in connection with the dependent variable, it indcates that some of the curves are not of proper shape. For instance, if a straight line is fitted to data having a decided non-linear trend, the errors plotted against the dependent variable will fall along a well defined U-shaped (or inverted U-shaped) curve.

Additional information may also be obtained by plotting the deviations against factors not included in the study. Relationships will sometimes become apparent which previously were obscured by the effect of the more important factors. The influence of such factors may account for seemingly abnormal cases and their inclusion would tend to reduce the mean and to a lesser extent the median deviation.

Frequency Distributions

Normal curves, such as those described above, form a basis for estimates of an *average value* for a group of items comprising a unit such

as has been utilized in developing the study. In many instances, however, it is necessary to know not only the average value but also the distribution of items about that average.

> Thus, the normal curves of the type of those shown in Fig. 4 serve as a basis for estimating the average message use of all subscribers to a given class of measured service in a given city. Additional curves are, however, required for estimating the distribution of subscribers by message use.

The basic principles governing the derivation of normal curves are the same whether these normals be concerned with averages or with distributions. The detailed methods involved are, however, quite different because of the inherent differences in the material. An average can be expressed in one arithmetic term which can be plotted against other factors. A distribution, on the other hand, is a complex entity which may itself be expressed as a curve but which obviously cannot be measured by an index to be plotted against other variables without losing sight of certain detailed characteristics of the distributions. The procedure and methods described above for deriving normal curves are modified somewhat in the derivation of normal distribution curves. Some of the methods which have been found advantageous for these analyses are described below.

The first step in the analysis is usually to plot the actual detail and cumulative distributions for each group of items and to compare the various distributions in order to determine points of similarity or difference. For purposes of comparison, percentage distributions are used, i.e., the per cent. of total items rather than the actual number occurring in each interval is plotted. With homogeneous material it will usually be found that when plotted to the same scales the detail frequency curves are all of the same general shape but differ in three primary characteristics.

1. The spread or extent of variation.
2. The location of the peak or point of maximum frequency.
3. The concentration of items in the peak interval.

These characteristics are, however, interrelated and to a certain extent related to the average.[7] Other things being equal, it might be expected that:

1. The greater the average the greater the spread.
2. The greater the spread the less the concentration at the peak.
3. The higher the peak the nearer it will fall to the average.

[7] Throughout this section the term "average" is used in the sense of arithmetic mean.

Since the average is of much importance in determining frequency curves, it will usually be found that differences will be reduced if the curves are plotted with each interval of the horizontal scale as a per cent. of the average instead of an actual value. For example:

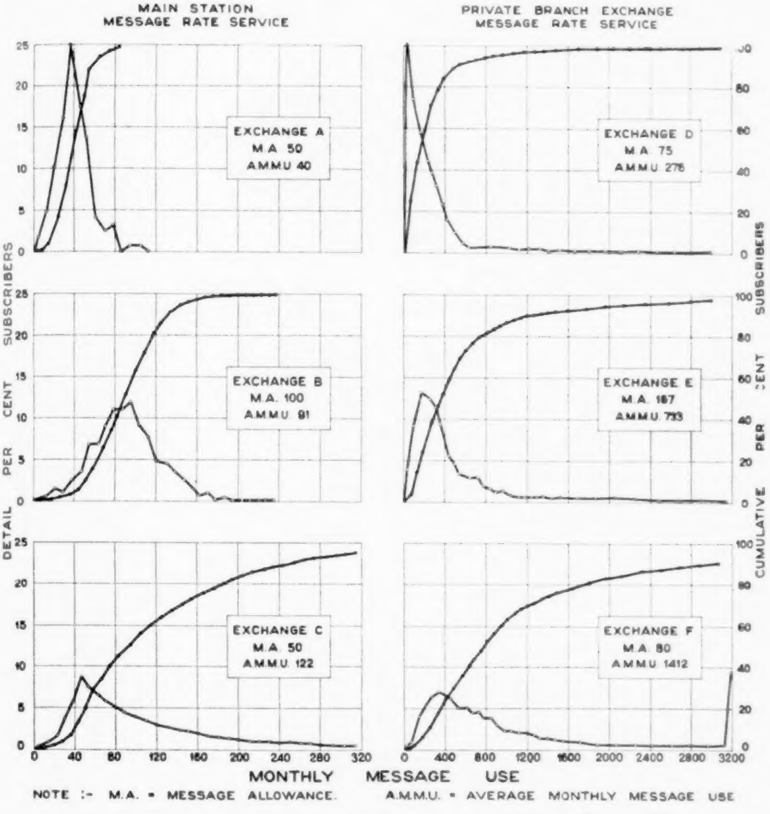


Fig. 5—Sample Distributions of Subscribers by Absolute Message Uses

Fig. 5 illustrates distributions of subscribers by message use plotted in terms of actual values for different classes of message rate service. On each chart the average message use is indicated. It will be noted that, in general, the greater the average message use, the greater the spread of the curves, the less the concentration at the peak interval and the less marked the correspondence of the peak with the average value. On Fig. 6 the frequency curves

have been replotted using instead of each actual message use interval the per cent. that the message use is to average message use. The result of this statistical process is to make the spread of the curves and the height of the peaks much more nearly uniform.

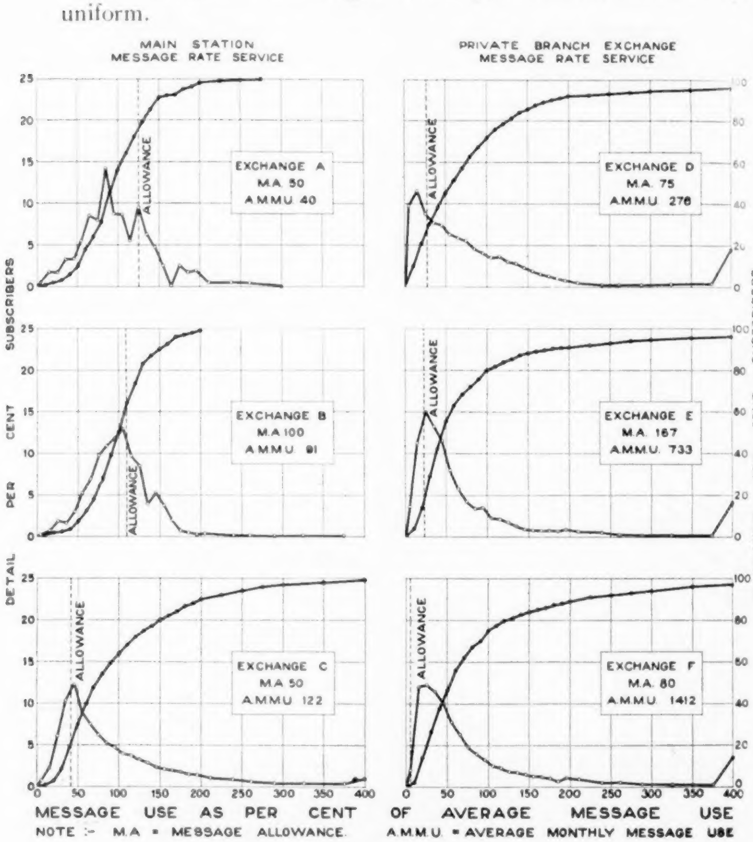


Fig. 6—Sample Distributions of Subscribers by Message Uses in Per Cent. of Average

In some instances, the frequency curves plotted with each interval expressed as per cent. of the average may be so similar for the different groups that satisfactory normals may be derived from this setup alone without including any other factor. This appears to be the case for the distribution of P. B. X. subscribers by message use as illustrated on Fig. 6. It is necessary, however, to test whether or not the full effect of the average on the distribution has been eliminated. This

may be done on a detail basis by plotting a series of charts showing the relation between the absolute amount of the average and the per cent. of cases falling within a given message use interval (expressed as per cent. of the average). On a cumulative basis the per cent. of cases falling below a given per cent. of the average is plotted against the average. For example:

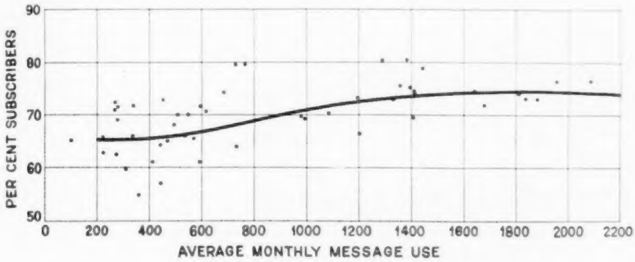


Fig. 7—Preliminary Chart of Cumulative Series

Fig. 7 shows the relationship between average message use and the per cent. of subscribers using less than 100 per cent. of the average message use for a given service classification. Each plotted point represents the reading from the cumulative curve for a different exchange. It is evident that the two factors vary together.

Curves similar in type to that shown in Fig. 7 are constructed on each of the charts of the detail and cumulative series. The curves of each series are smoothed by cross-sectioning and developed into consistent and reasonable families.

In connection with the smoothing of the cumulative series a method described below has been found useful. This method can be used with any setup of three variables but is simplest in setups of a cumulative type which have no maxima or minima within the limits of the curves. To simplify the explanation of the method the cumulative distribution of certain subscribers by message use is referred to as an example.

Fig. 8 shows preliminary curves representing the relationship between average message use and the per cent. of subscribers using less than the various per cents. of the average message use from 10 per cent. to 500 per cent. of the average. These curves are derived from a series of charts similar to Fig. 7 for different message uses. Cross sections of the family of curves of Fig. 8 give a series of cumu-

lative curves. The successive curves of this cumulative series have been plotted in Fig. 9 at regular intervals apart, the intervals being the same distance as the average monthly message use scale of Fig. 8. The horizontal scale of Fig. 9 used in plotting these cumulative distributions must therefore be movable so as to apply in turn to each

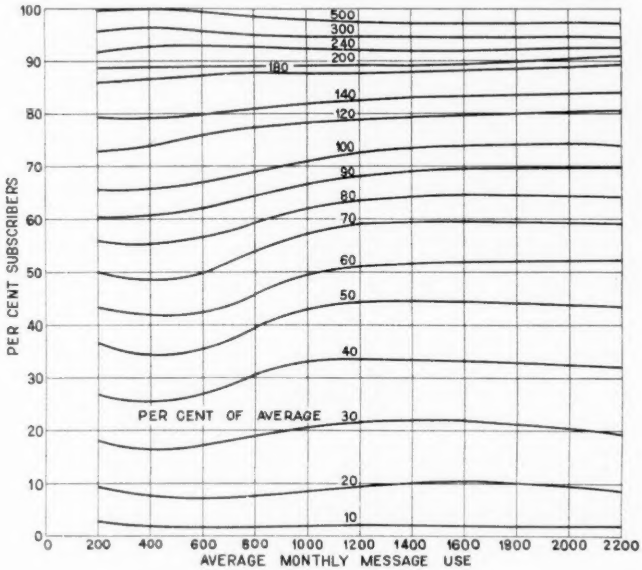


Fig. 8—Preliminary Family of Cumulative Curves

of the cumulative curves. With the cumulative curves plotted, the family of curves on Fig. 8 have been drawn in on Fig. 9, the curves representing the various message uses being exactly the same as those of Fig. 8 except that the method of drawing the cumulative curves has automatically shifted successive curves of Fig. 8 further and further to the right. It follows from the methods which have been used in constructing Fig. 9 that any given cumulative curve must intersect each curve of the other family somewhere on the vertical line corresponding to the message use (expressed as per cent. of the average message use) represented by that curve. For instance, the cumulative curve for 1000 average message use (indicated by A) must intersect the curve representing 30 per cent. of the average message use on the vertical line corresponding to a co-

ordinate of 30 on the horizontal movable scale when the zero point of the horizontal movable scale falls at 1000 average message use on the fixed horizontal scale. The point of intersection described in this illustration is indicated on Fig. 9 by P. This characteristic (inter-

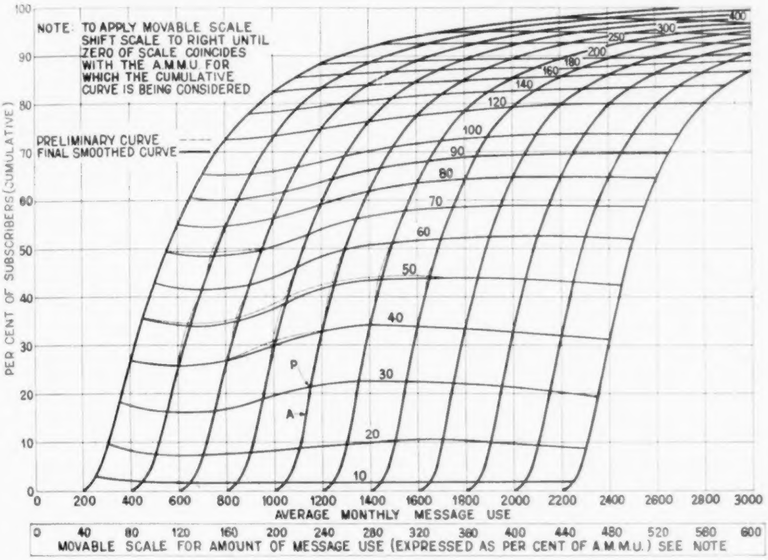


Fig. 9—Simultaneous Smoothing

sections falling on given vertical lines) forms the basis for smoothing the two families of curves simultaneously. A point of intersection may be shifted vertically but cannot be shifted horizontally since it must fall somewhere on a definite vertical line. Dashed lines (- - - -) on Fig. 9 indicate the manner in which a few of the points have been shifted in smoothing.

A family of cumulative curves may appear easier to smooth than the corresponding family of detail curves. On the other hand, the detail curves give, in some respects, a more vivid picture of the outstanding characteristics of the distributions than do the cumulative, and certain important characteristics of the distributions may be more easily studied on a detail basis.

It is important, therefore, that both series be taken into account in deriving final normal distribution curves. For the detail series, charts are plotted showing the relationship between the amount of

the average and the per cent. of cases falling in a particular interval (expressed as per cent. of the average). Fig. 10 is such a chart for the interval 80–90 per cent. of the average message use.

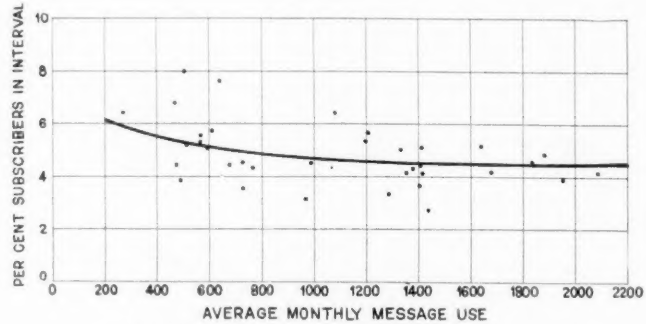


Fig. 10—Preliminary Chart of Detail Series

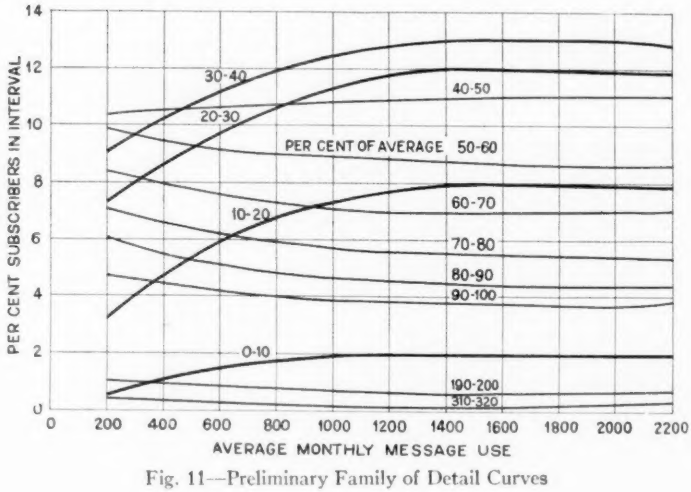


Fig. 11—Preliminary Family of Detail Curves

Cross-sections of a family of curves such as those on Fig. 11 give a series of detail frequency curves. Further smoothing may be facilitated by a study of these curves. As an aid in this process of smoothing it is desirable to determine the normal location of the peaks and the normal proportion of cases occurring in the peak interval, as these are important characteristics of such curves. These normal values may be determined by plotting these factors against

the average as shown on Figs. 12 and 13. For this data, it is noted that, on an absolute message use interval basis, the greater the average the greater the abscissa of the peak value. However, with an increase in the average, the abscissa of the peak interval on an absolute basis

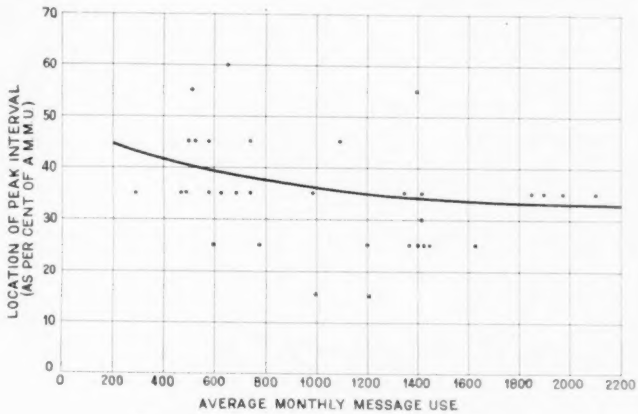


Fig. 12—Determination of Normal Peaks

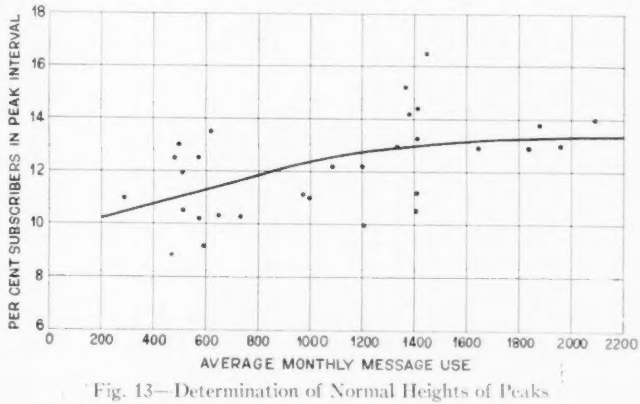


Fig. 13—Determination of Normal Heights of Peaks

does not increase as rapidly as does the average, because large users increase their usage relatively more than small users. Therefore, when the intervals are plotted in terms of per cent. of the average, it will be found that the greater the average the less the abscissa of the peak. For the same reasons, the peak interval expressed as per cent. of the average becomes relatively wider as the average increases, and the height of the peak increases as illustrated on Fig. 13. With

the location and height of the peaks normally determined, the process of constructing preliminary curves for the various intervals of the detail series is, in many cases, considerably simplified. These preliminary curves are then cross-sectioned and smoothed into a consistent family.

Finally the smooth curves of the cumulative and detail series are checked with each other and averages are computed from these curves as a check against the assumed average. When the minor discrepancies disclosed by these checks have been corrected, normal curves are plotted and comparisons are made with the actual distributions. Further adjustments may then be necessary.

Additional Factors in Deriving Normal Distribution Curves

The smoothing processes described above give a series of normal [8] distribution curves taking into account completely the effect of the amount of the average upon the distribution. In some cases, however, it will be found that some outside factor has also a decided effect upon the distribution.

When the effect of an outside factor is apparent it may be necessary to derive a series of normal distribution curves, each curve corresponding to a constant value of the factor under consideration. If this is done, the curves are smoothed by cross-sectioning and the various other methods described above so as to form a consistent and reasonable family. The type of the final family derived will, however, depend largely upon the character of the relationships developed during the smoothing process. For instance, in the case of main station message rate service, a series of distribution curves was plotted, one for each message allowance. In the course of smoothing these curves it seemed reasonable that there might be a relationship between the type of distribution and the proportional relationship of average message use to message allowance. That is, with an annual message allowance of 600 and an average annual message use of 400 the distribution of subscribers by amount of message use might be similar, on a proportional basis, to the distribution of subscribers under an annual message allowance of 900 with an average annual message use of 600; or under an annual message allowance of 1,200 with an average annual message use of 800. This idea was tested by use of the various sets of normals which had been derived for the different message allowances and was found to hold so closely that this pro-

[8] See Note 6.

portion factor (ratio of the average message use to the message allowance) might be used in deriving a revised setup of normal distribution curves.

In certain cases it may be found that some expedient such as that described above may be used to eliminate or take account of the effect of an outside factor. Whether this is done or a separate set of curves is derived for different values of that factor, the process of deriving the detail and cumulative distribution curves would in general be the same as that described above.

Some of the processes involved in studying averages and distributions of subscribers by message use have been described because they are typical and illustrate what have been found to be satisfactory methods of analysis for problems of this type. It is clearly impossible, however, to set up any rigid methods for such studies. Any economic problem which permits of analysis by these methods must be treated in the manner best suited to the data available, the purposes of the study, the degrees of accuracy necessary, etc. Where these methods can logically be employed the results obtained, an important part of which are the background and sidelights, on the problem, disclosed during the process of building up the normal relationships, will generally be found superior to those obtained through the use of more rigorous methods.

Application

Before final results are obtained, there will naturally be developed by those concerned in the study a very definite conception of the field of their usefulness and their limitations. It is important that a knowledge of these limitations be extended to those who may have occasion to use the results. Given a set of smooth curves from which quantitative estimates can be made, there is a great temptation to make estimates under any and all circumstances, and often to give such estimates an undue appearance of accuracy. The final results are merely the general expression of the information contained in the original data logically developed according to the knowledge and judgment of the investigator. It is always necessary in applying such results to consider the effect of special and local conditions. Where it is known that actual conditions in a specific case are far from normal, it is often possible to estimate the effect of a proposed change by applying differences based on the normal experience.

Care must also be taken in extrapolation estimates, i.e., estimates where the value of one or more factors lies beyond the limits of the original data. Such estimates, of course, are always subject to con-

siderable error. In other cases it may happen that some part of the data necessary for making a complete estimate is not available. It may be practicable, however, to approximate the required information and make a rough estimate which may be more accurate than the alternative of basing the estimate upon less complete analysis.

In applying the results of such analyses, satisfactory conclusions can be reached only if due consideration is given to the following points:

1. The quantitative readings from the normal curves.
2. All the qualitative relationships developed in the course of the analysis.
3. Any additional data available for the particular case or cases in question.
4. Any peculiar special conditions known to exist for that case or which probably exist because of comparison with similar cases.
5. Changes affecting general levels since the date of the study.

It follows that the making of such estimates cannot be left in inexperienced hands any more than can the progress of the original study. Good judgment and a complete knowledge of the problem are of paramount importance both in making the general analysis and in the application of results to specific problems.

To those accustomed to working in the more exact fields of physics, chemistry, etc., it will undoubtedly appear that the methods described above may be inexact and unsatisfactory. Undoubtedly, the average errors of estimate are considerably greater than would be allowed in fields where more exact data are obtainable. Yet the reason for this lies rather in the material itself than in the methods of dealing with it. An economic quantity is extremely complex and difficult to estimate because it is usually dependent upon the action of hundreds or thousands of individuals each one of whom is influenced by individual needs and desires which at best can only be partially measured by such quantitative factors as reflect variations in these needs or desires. Estimates of such quantities are necessarily subject to a relatively high degree of error if comparisons are made with the fields of physical science. Yet such estimates must be made and the problem is to make them as accurately as practicable. Judged from this standpoint, experience indicates that such analyses are an important aid in connection with certain phases of many economic problems.

Permalloy, A New Magnetic Material of Very High Permeability

By H. D. ARNOLD and G. W. ELMEN

SYNOPSIS: The magnetic alloy described in this paper is a composition of about 78.5% nickel and 21.5% iron and at magnetizing fields in the neighborhood of .04 gauss and with proper treatment has a permeability running as high as 90,000. This is about 200 times as great as the permeability of the best iron for these low magnetizing fields. This high permeability is attendant upon proper heat treatment and also upon other factors among which is freedom from elastic strain. The presence of other elements than iron or nickel and specially carbon, reduces the permeability, but slight variations in heat treatment produce large changes compared with those due to small quantities of impurities.

So far as discovered, other physical properties show no peculiarities at the composition which brings out the remarkable magnetic properties of permalloy. The equilibrium diagram, electric conductivity, crystal structure, mean spacing between adjacent atom centers and density are among the physical properties which have been studied.

To the engineer in electrical communication the development of permalloy is very significant. It assures a revolutionary change in submarine cable construction and operation and promises equally important advances in other fields.—*Editor.*

SOME time ago it was discovered in the Bell System laboratory[1] that certain nickel-iron alloys, when properly heat-treated, possess remarkable magnetic properties. These properties are developed in alloys which contain more than 30 per cent of nickel and which have the face-centered cubic arrangement characteristic of nickel crystals, rather than the body-centered structure characteristic of iron. The entire range above 30 per cent nickel exhibits these properties to some degree and offers new possibilities to those interested in magnetic materials. The most startling results, however, are obtained with alloys of approximately 80 per cent nickel and 20 per cent iron, whose permeabilities at small field strengths are many times greater than any hitherto known. To alloys of this approximate composition we have given the name "permalloy". The development of permalloy has assured us a revolutionary change in submarine cable construction and operation, and promises equally important advances in other fields of usefulness. It also presents questions of great interest to the scientist, and emphasizes again the meagreness of our fundamental information about ferromagnetism. The present paper is intended to give a general discussion of the preparation and testing of permalloy, with sufficient detail to indicate its unusual characteristics. Detailed statements of numerical results are reserved for publication in separate articles dealing with specific properties.[2]

[1] British Patent No. 188,688.

[2] L. W. McKeehan, The Crystal Structure of Iron-Nickel Alloys, Phys. Rev. (2), 21, (1923).

In making permalloy we use the purest commercial nickel and Armco iron. Our samples for laboratory study are prepared by melting these metals in a silica crucible, using a Northrup high-frequency induction furnace. The particular furnace which we use will conveniently melt a charge of about six pounds. An analysis typical of the resulting billets is as follows:

Ni	78.23
Fe	21.35
C	.04
Si	.03
P	trace
S	.035
Mu	.22
Co	.37
Cu	.10

The presence of other elements than nickel and iron is of course to be expected after any practical method of preparation. To determine their effects, samples were prepared in which the usual impurities were present in various proportions. It was found that their presence does affect the permeability of the alloys and that carbon is especially harmful. Since, however, the variations produced by slight changes in heat-treatment are very large compared with those due to small quantities of impurities we have found it unnecessary for most purposes to require higher purity than that indicated in the analysis above given.

In our laboratory studies we have made it a practice to reduce the billets through the forms of rod and wire to tape 3.2 mm. wide and 0.15 mm. thick. Accordingly test samples are available in a variety of forms and conditions. Thin narrow tape is particularly adapted to use in experiments involving heat-treatment, since it possesses a high ratio of area to volume and is easy to manipulate. Fortunately the entire nickel-iron series can be mechanically worked if sufficient care is exercised and we have thus been able to use samples of the same size, shape, and mechanical condition in all measurements upon which we have based comparisons between alloys. This practice has also made possible strictly comparable micrographic studies throughout the series.

Permeability is the magnetic characteristic of permalloy in which we first became interested and we have used its numerical value as an index in establishing the effects of mechanical and thermal treatments. Most of the measurements have been made in a ring permeameter of special design. The ring sample is prepared by winding twenty or more turns of tape around a disk about three inches in diameter. The disk is then removed leaving the material in the form of a spirally laminated ring with a rectangular cross-section approximately 3.2 mm. by 6 mm. A single massive copper conductor is linked with this ring, and constitutes also the secondary of a transformer whose primary winding forms one arm of an inductance bridge. From the bridge measurements, and the dimensions of the ring the permeability of the latter may readily be computed. For most of the measurements 112-cycle alternating current has been employed, permitting the use of telephone receivers in adjusting the balance of the bridge. The ring is sufficiently well laminated so that no serious troubles are introduced at this frequency by eddy currents. This fact was verified by making a number of permeability determinations at alternating current frequencies both above and below that chosen for routine use, and also by comparing the results of ring permeameter tests with those of ballistic tests on specially wound ring samples. The bridge method is particularly well adapted to the measurement of permeability in very weak magnetic fields since amplifiers may readily be used to increase the delicacy of the bridge adjustment to almost any degree desired. As a matter of convenience we have usually included in our test program measurements with fields of 0.002, 0.003, and 0.010 gauss, and on the graph of permeability against magnetizing field strength the straight line through these points has been extended to field strength zero. We have called the permeability read from the graph at this point the "initial permeability" of the sample.

The form of permeameter used is especially adapted to making measurements quickly and with minimum handling of the sample, since it makes use of but a single magnetizing turn. The ring is laid on suitable insulating supports in an annular copper trough, and placing the copper cover on this trough completes the electrical circuit. In a modified instrument, the "hot permeameter", provided with a heating device, permeabilities may be measured from liquid air temperatures up to about 1000°C. without altering the position of the sample.

The heat-treatment of permalloy is of the utmost importance. To develop its maximum initial permeability it must be cooled not only through the proper temperature ranges, but also at the proper rates.

It is obvious that only a small part of any sample can be given the most favorable treatment, since the interior portions of the sample cool at rates which are dependent upon the geometrical configuration and thermal properties of the material and are only indirectly under the control of the experimenter. For these reasons each shape and size of sample will have its own best heat-treatment and it is obviously difficult to establish the correct heat-treatment for a small element of volume, characteristic of permalloy as a material. By the use of thin tape, however, we secure fairly uniform treatment of the whole volume so long as the cooling is not too rapid, and fortunately the best cooling rate is not much different from the normal cooling rate of the tape in the open air. It has been found that temperature changes below 300°C. have very little effect upon the resultant properties of permalloy, but the rate of cooling from just above the magnetic transformation temperature down to about 300°C. is a controlling factor. By a long series of experiments a heat-treatment has been established which is especially well adapted to the permalloy test rings already described. They are first heated at about 900°C. for an hour and allowed to cool slowly, being protected from oxidation throughout these processes. They are then reheated to 600°C., quickly removed from the furnace and laid upon a copper plate which is at room temperature.

Not only does each size and shape of sample require its own special heat-treatment, but samples differing only in composition also differ in their most suitable heat-treatments. In our investigation of the nickel-iron series we have not, however, attempted to determine the best heat-treatment for ring samples of each of the many alloys studied. By careful exploration we located the region about 80 per cent nickel, 20 per cent iron as the one promising the highest initial permeability and established the best heat-treatment for this composition. Keeping this treatment unchanged we then relocated the best composition, finding it to be at about 78.5 per cent nickel, 21.5 per cent iron. There is a maximum temperature in the equilibrium diagram for this binary at about 70 per cent nickel,[3] and it was natural to suspect that the maximum in initial permeability which we had found at 78.5 nickel might be displaced to 70 nickel by proper treatment. The 70 per cent nickel alloy was accordingly subjected to a great variety of heat-treatments, but no method was found capable of producing in it an initial permeability as high as that readily obtainable in the 78.5 per cent nickel alloy.

Fig. 1 shows the general way in which initial permeability has been found to vary throughout the nickel-iron series when the heat-treat-

[3]Bureau of Standards Circular No. 58, April 4, 1916.

ment determined as best for the 80 per cent nickel alloy is used. It is obvious from what has been said above that too much weight must not be given to the actual values recorded at any composition. Had the best heat-treatment been determined for each sample the curve might have been altered considerably in detail, particularly outside the permalloy range. We believe, however, that its general form is

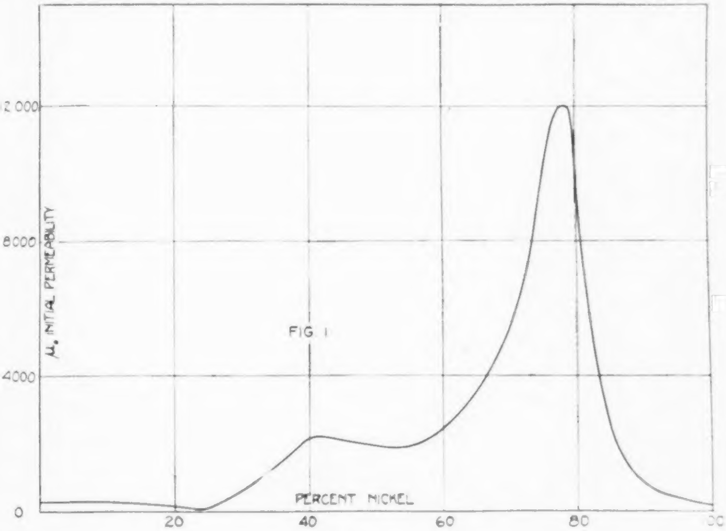


Fig. 1

approximately correct. Alloys were made at 5 per cent steps throughout the range except in the vicinity of 80 per cent nickel where a great number of slightly different compositions were investigated. The chemical analysis, rather than the intended composition, was used in every case, although the difference was never considerable.

The largest value of initial permeability for permalloy at room temperature which we have so far found in the ring permeameter is about 13000, more than 30 times the corresponding value for the best soft iron. How extraordinary this is may be appreciated by considering that this material, although it has a saturation value of magnetic intensity comparable with that of iron, approaches magnetic saturation in the earth's field. Unusual caution must therefore be exercised in measuring the properties of permalloy to protect the sample from the influence of stray magnetic fields. Fig. 2 shows, to

different scales, the values of initial permeability in similar ring samples of permalloy and of annealed armco iron, and small portions of the corresponding μ-H curves from which these were obtained.

We have measured the magnetization of permalloy at saturation and find that it is not sensitive to heat-treatment. The saturation values of magnetization per gramatom are known to vary almost

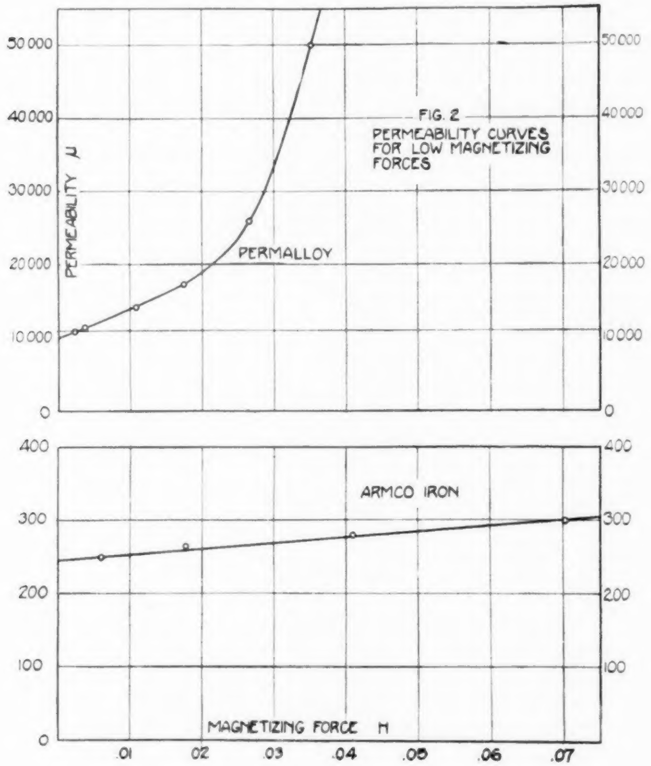


Fig. 2

linearly with composition throughout the nickel-iron series, from 222 for iron to 59 for nickel.[4] The value 84 which we have found for the 78.5 per cent nickel alloy is therefore not abnormal.

The magnetic characteristics of heat-treated ring samples of the same alloy have also been determined through a wider range of field

[4]P. Weiss, Faraday Society Trans. 8, 149-156 (1912).

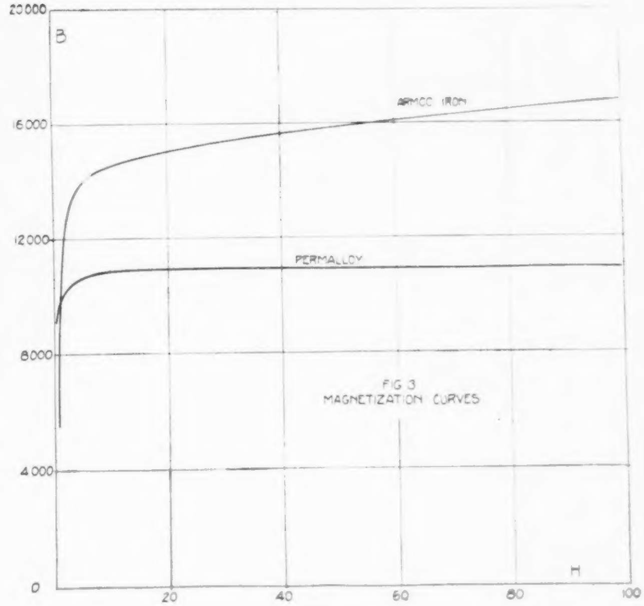

20000
B
ARMCO IRON
16000
12000
PERMALLOY
8000
FIG 3
MAGNETIZATION CURVES
4000
0
H
20
40
60
80
100

Fig. 3

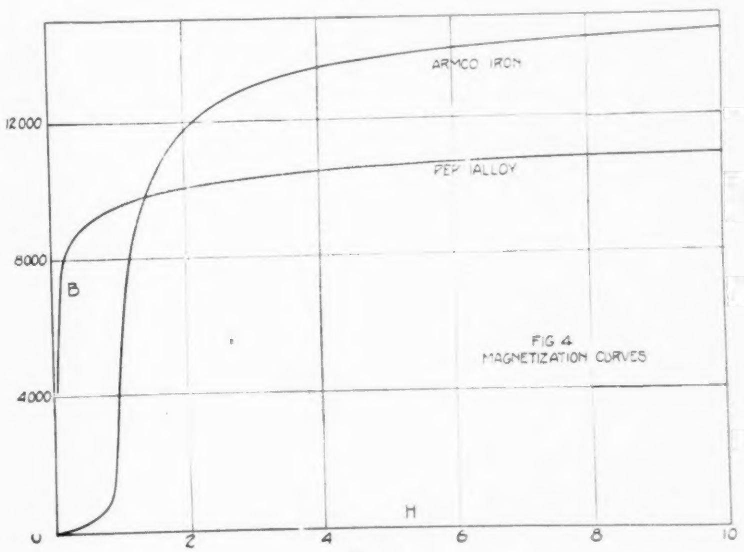

ARMCO IRON
12000
PERMALLOY
8000
B
FIG 4
MAGNETIZATION CURVES
4000
H
0
2
4
6
8
10

Fig. 4

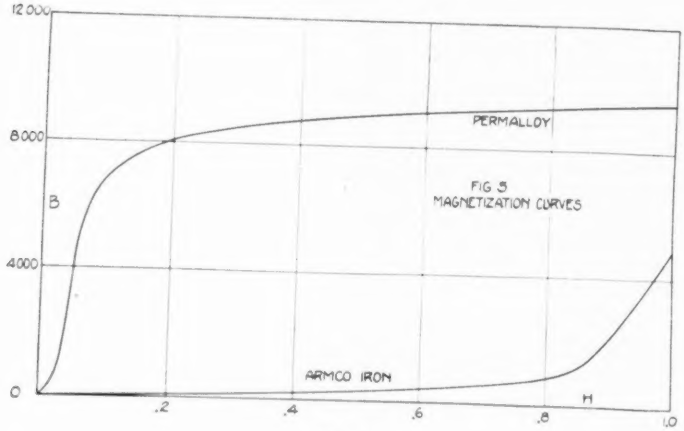

12000
8000
4000
0
B
PERMALLOY
FIG 5
MAGNETIZATION CURVES
ARMCO IRON
H
.2
.4
.6
.8
1.0

Fig. 5

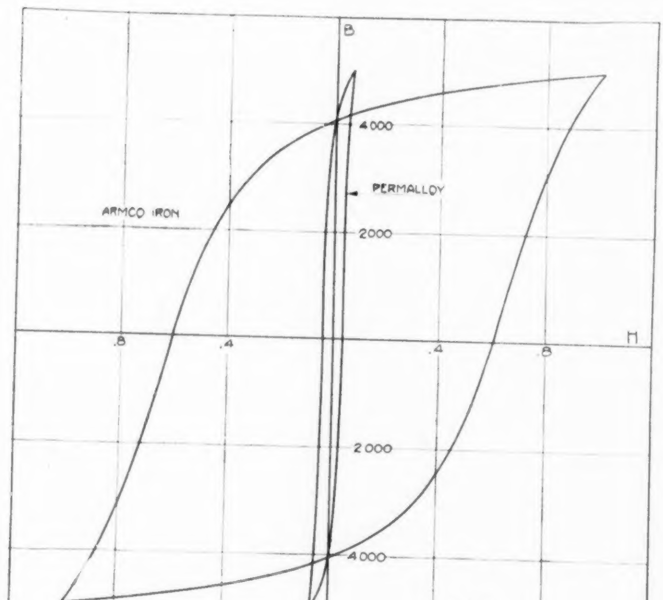

B
4000
PERMALLOY
ARMCO IRON
2000
H
.8
.4
.4
.8
2000
4000
FIG 6

Fig. 6

strengths by ballistic methods. Figs. 3, 4, and 5 show B-H curves for such a sample of permalloy and for a sample of annealed armco iron. From Fig. 5 is apparent the enormous susceptibility of the former material in the weak fields so important in communication engineering. Fig. 6 shows for the same two materials hysteresis loops carried to a maximum induction of 5000 maxwells. The area of the permalloy loop is only one sixteenth that of the loop for soft iron. Fig. 7 shows

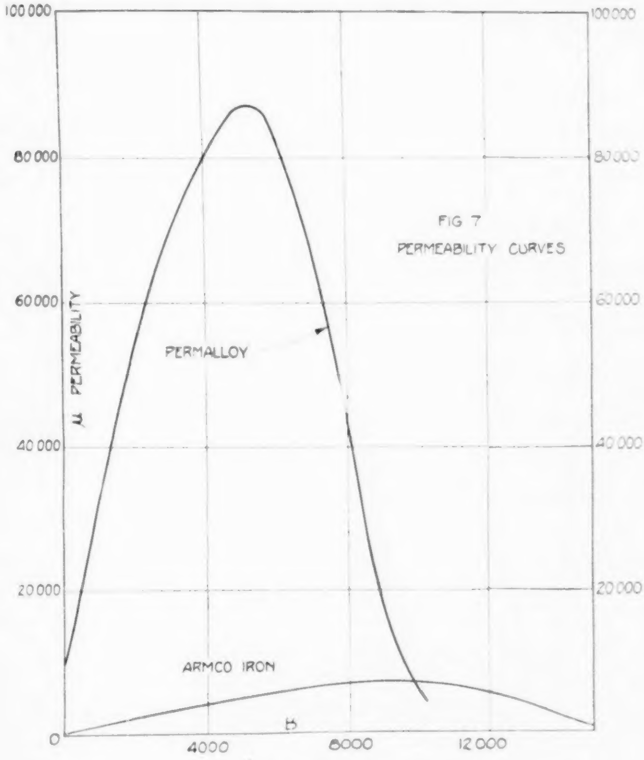


Fig. 7

the μ-B curves for these materials. The maximum permeability here shown, $\mu = 87000$, which is not exceptionally high for permalloy largely exceeds the highest values obtainable in silicon steel[5] and of course occurs at a much lower flux density.

[5]T. D. Yensen, U. S. Patent 1,358,810.

Early in the investigations it was found that heat-treated permalloy is sensitive to strain, and the routine measurements were so conducted as to avoid this disturbing effect. Separate investigations of the effects of strain upon permeability and electrical conductivity in straight samples, and of the converse effects of magnetization upon dimensions and conductivity were also undertaken. While these studies are not yet complete it can be stated that all these effects are large in comparison with the corresponding effects in hitherto available magnetic materials. So long as the elastic limit of the material is not exceeded the effects due to strain are reproducible and disappear when the strain is relieved. The effects of magnetization, however, show the expected hysteretic properties. As an example of the magnitude of the effects producible it may be stated that between its value in the unstrained condition and about one-tenth that value the initial permeability of a heat-treated strip of certain of these materials can, by the mere variation of strain, be adjusted to any value we may for the moment desire. The range through which the conductivity can similarly be adjusted by strain is much narrower, the maximum reduction being about 2 per cent, which, however, is a large effect compared with that found in other metals.

The effect of magnetization in reducing conductivity is as much as 2 per cent for fields of the order of one gauss. This makes it easy, for example, to measure the earth's magnetic field to within about 1 per cent by finding the strength of the opposing field necessary to give a permalloy strip its maximum conductivity. It will be noted that the conductivity change which we have mentioned as attainable by magnetization is the same as that attainable by elastic strain. This is no mere coincidence, for we find that the maximum change due to either cause alone is not further increased by superposition of the other, although the effects of small tensions and magnetizing fields are additive. This suggests, of course, that both causes ultimately produce the same change in the mechanism responsible for conduction.

Since the effect of tension upon permeability is in some of these cases so marked it seemed surprising that the only reported study[6] of the converse effect, that is of magnetostriction, indicated a zero value within the permalloy range. It appeared advisable therefore to study the magnetostriction of the series of alloys here available. Preliminary results indicate that under usual conditions of experiment, heat-treated 78.5 per cent nickel alloy exhibits larger magnetostriction than does iron.

[6] K. Honda and K. Kido, Tohoku Univ. Sci. Rep., 9, 221-232, (1920). It should be noted, however, that their alloys had received different treatments than ours.

With the remarkable ferromagnetic behavior of permalloy in mind one naturally looks for analogous peculiarities in its other properties. As has been shown, however, the equilibrium diagram does not point accurately to the composition exhibiting highest initial permeability. The conductivity curve is even less indicative of a peculiarity at this point, its minimum lying at about 35 per cent nickel. The crystal structure is that of nickel and its type does not change until the nickel content is made less than 35 per cent. Even the mean spacing between adjacent atom-centers, and with it the density, varies continuously throughout the entire range. Our experience in working these alloys also indicates that the series has no mechanical peculiarities at or near 80 per cent nickel. Not only do these characteristics indicate no abnormality as the nickel content is increased beyond 70 per cent, but, what is more surprising they are little affected by the heat-treatments which so profoundly change the magnetic properties. So far as has been determined, therefore, it is only in connection with its magnetic properties that permalloy is unusual.

To the engineer the discovery of permalloy means the realization of plans long impossible of accomplishment for lack of a suitable material. For the scientist the principal interest in these materials may well lie in the large response of their magnetic properties to simple external controls. Without alteration of composition these properties may be adjusted through extraordinary ranges by strain, by magnetization, or by heat-treatment. This allows a more definite study of the way in which these factors are related to magnetic properties than has been possible with materials hitherto available in which their effects are comparatively small and may be associated with complicated and irreversible changes in other properties. The behavior of permalloy demonstrates that ferromagnetism is associated with material structure in a different way than are the ordinary physical and chemical properties and its extreme sensitiveness to control gives us a powerful method for use in magnetic investigations.

Telephone Equipment for Long Cable Circuits[1]

By CHARLES S. DEMAREST

SYNOPSIS: Some of the important developments contemplated in the apparatus and equipment for long toll cable circuits are described. The large number of equipment units per station in the cable plant and the greater number of stations in a given length of cable than in an open-wire system have made the economic importance of the equipment design such that a comprehensive program of development, affecting many types of equipment, has been undertaken. The outstanding features of some of the more important of these, including the telephone repeater equipment, test board equipment and signaling equipment, are described. The necessity for compactness in the dimensions of equipment units, uniformity in assembly arrangements, and simplicity in design, together with the need of careful correlation of the electrical and mechanical requirements, are emphasized. The methods proposed for meeting these requirements generally, are described.

INTRODUCTION

THE use of lead covered cables in place of bare copper wires for long distance telephone lines has been an important development and much interesting information on this subject has already been presented to the Institute. The engineering and construction features involved in a cable system of this sort were described by Mr. Pilliod[2] in his article on the Philadelphia-Pittsburgh Section of the New York-Chicago cable, while the transmission characteristics of such a system were brought out in the recent paper by Mr. Clark.[3] It is the purpose of the present paper to deal with some of the important developments in apparatus and equipment which are contemplated for the cable plant.

A cable system requires repeater stations at more frequent intervals throughout its length than an open-wire line, because of the much smaller gauge conductors which it employs and the increased electrical capacity due to closer proximity of the wires. Consequently, in such a system, a greater proportion of the plant investment is represented by the equipment within the offices than is the case with open-wire construction. Furthermore, the number of equipment units per station in a cable system is ordinarily much larger than in an open-wire office, due to the fact that the chief advantages in the use of long cable circuits, in place of open-wire construction, have occurred on routes carrying heavy traffic where many circuits are

[1] Presented before A. I. E. E., June 27, 1923.

[2] Journal of the A. I. E. E. for August, 1922; and *Bell System Technical Journal*, July, 1922.

[3] Transactions of A. I. E. E., Vol. 38, part 2, p. 1287; and *Bell System Technical Journal*, January, 1923.

needed. Thus, the requirements of the cable plant have been such as to emphasize the economic importance of the equipment design.

To meet these requirements it has been necessary to undertake a comprehensive plan of development affecting many types of equipment. This has involved careful consideration of both the electrical arrangements and the mechanical design, which are being closely coordinated with the purpose that both should contribute to the highest degree of efficiency in the functioning of the system.

It is not possible in this paper to give many details concerning these developments, but it is desired to present some of the principal features as applied to typical cases. Among the more important of these are the telephone repeater equipment, the testboard equipment and the signaling equipment. These are closely associated with each other in their operation, as well as in their physical location, and it has been necessary, in the design of all units to have due regard to the system as a whole.

Telephone Repeater Equipment

The function of the telephone repeater as an amplifier in long distance lines is well known. The telephone repeater in its present form has been the chief factor in making long distance cable telephony practicable, and it is probable that the developments in connection with telephone repeaters have been among the most rapid and comprehensive of any in the toll equipment. It will, therefore, be very interesting to note, in the illustrations which follow, the principal

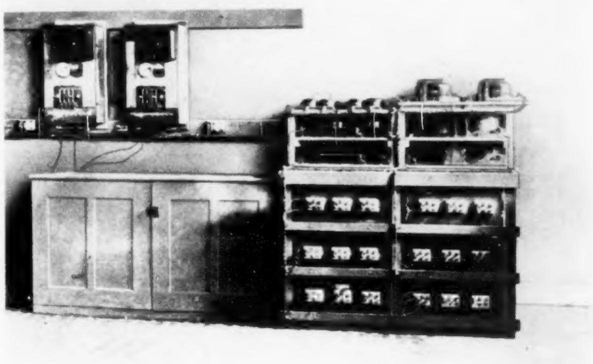


Fig. 1—Box Type Repeater Installation

steps which have been taken in working out the form of the equipment to the degree of efficiency now required for the cable plant.

In Fig. 1 is shown one of the original forms of repeaters, a number of which were installed as early as 1914. In this case the repeater apparatus was assembled in boxes designed to mount on the wall, each box containing a one-way amplifier. Two such one-way units would

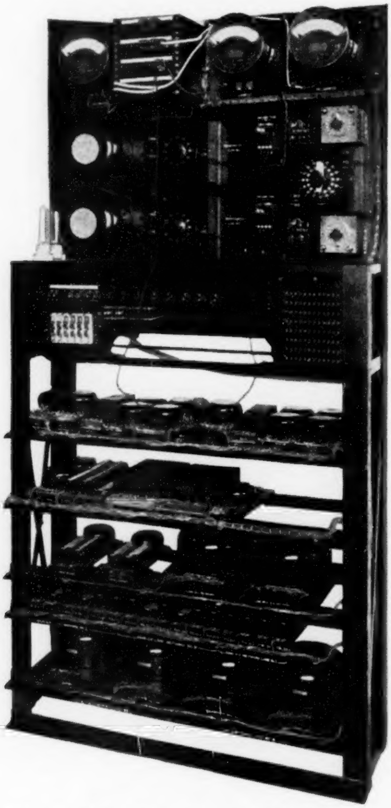


Fig. 2—Wide Rack Through Line Repeater

be required to form what is now known as a two-way, two element repeater. Although the particular amplifiers shown in the illustration were actually used for one-way operation only, they are typical in their general form, of the units employed for two-way operation. The balancing networks, associated coils, etc., were

mounted on separate racks while the plate batteries for the vacuum tubes were mounted in an adjacent enclosed cabinet. The floor space required for a two-way repeater employing this type of apparatus was in the neighborhood of 15 square feet per unit, including the usual allowances for aisle space and associated apparatus.

Fig. 2 shows the first type of vacuum tube telephone repeater designed for commercial manufacture. The particular unit shown is a single "through line" set, that is, one which remains connected

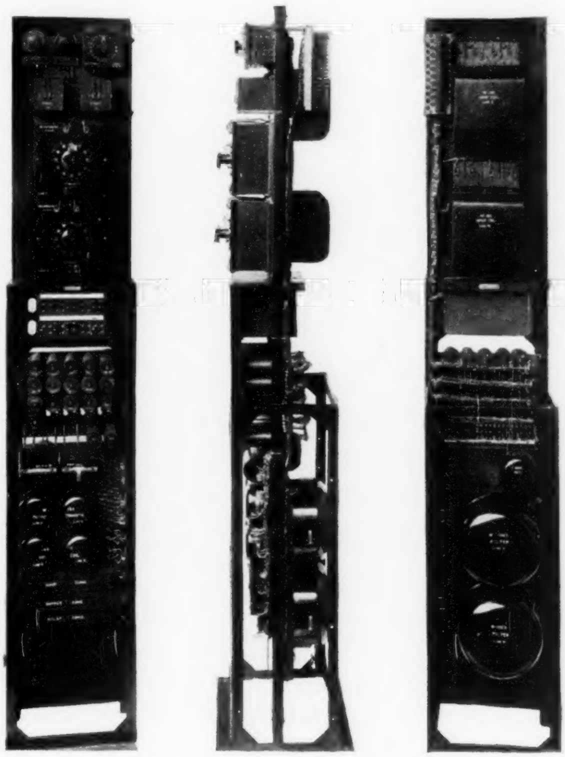


Fig. 3—Standard Through Line Repeater for Open-Wire Use

to a through circuit at all times instead of being used at a switching point to establish built-up connections at the will of the operator, as a "cord circuit" repeater is used. The "cord circuit" repeater of the same type was similar in arrangement and dimensions to the "through line" repeater shown. In this type of repeater, all of the

apparatus for a two-way circuit was mounted together on one rack. The testing equipment and signaling apparatus were duplicated in each repeater set. This apparatus, as well as the balancing networks and other miscellaneous apparatus were mounted on the same rack as the repeater. A unit of this type required about 10 square feet of floor space.

Fig. 3 shows the type of "through line" set which was standardized in 1917 for use on open-wire lines. Fig. 4 shows a group of cord circuit

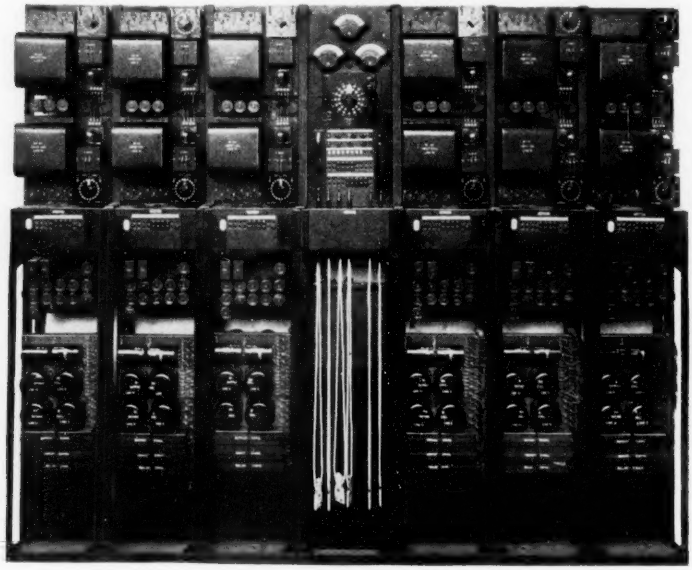


Fig. 4—Group of Six Cord Circuit Repeaters and Associated Testing Unit

sets of the same type of design, together with the testing unit. This form of set was a great improvement over the earlier types. It employed, however, many of the same large types of individual pieces of apparatus as were used in the former sets and the testing equipment, which was mounted on the middle rack of the group, was required to be duplicated for each group of 6 repeaters. This type of set has been used in many of the smaller installations, but it has not met the requirements for large cable installations. The average floor space area required was about 6 square feet per set.

Fig. 5 shows the proposed assembly of the type of telephone repeater which is now being developed for all classes of installations

employing the two-way, two-element circuit. Fig. 6 shows the general arrangement proposed for a group of sets of this type in a large installation, as in a cable office. This set is expected to have

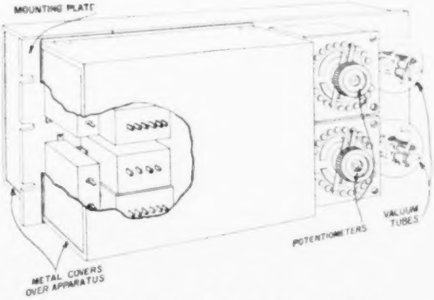


Fig. 5—Typical Assembly of Panel Mounted Repeater Set

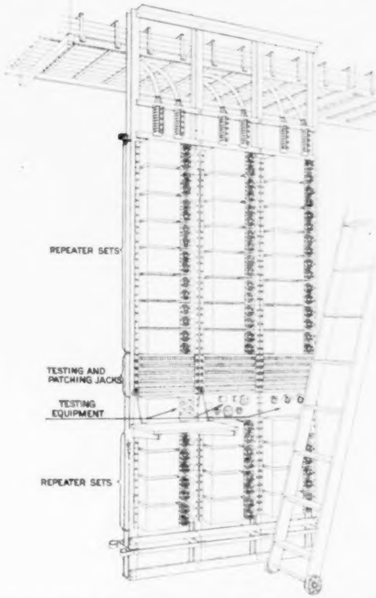


Fig. 6—Group of Panel Mounted Repeaters as Arranged in a Large Installation

many advantages adapting it particularly to cable installations. When mounted as shown in Fig. 6 it will occupy but 1.5 square feet of floor space per unit. Fig. 7 shows how this set may be arranged

in small installations where it may be desired to be mounted on a low rack.

By the uniform use of these general mounting arrangements for all of the new repeater equipment, including the accessory apparatus as well as the repeater sets themselves, it will be possible, where desired, to serve a large number of repeaters with a small amount of testing

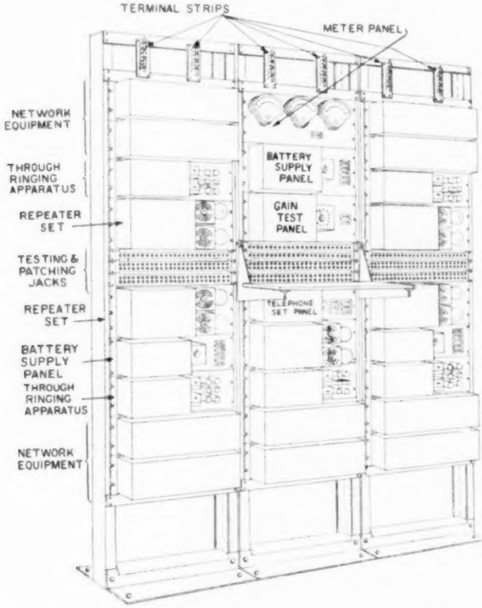


Fig. 7—Panel Mounted Repeaters as Arranged in a Small Installation

equipment. For example, in the case of the voltmeters and ammeters required, it will be possible to employ but one meter panel for as many as 120 repeaters. Thus, an economy in equipment as well as a saving in space will be effected.

Fig. 8 shows how some of the principal features of this proposed type of set, which distinguish it from the earlier types of repeaters, are related to the circuit arrangement, as well as to the mechanical design. Previously, the apparatus which it is now proposed to mount in distinct groups on separate panels, as indicated in this diagram, was assembled together in one repeater unit. Several types of sets were accordingly necessary to meet the various field conditions. This is to be avoided in the new design by separating from the basic re-

peater unit such apparatus as may be required to be different under different conditions of use. For example, the basic repeater unit in the new repeater is to be the same for both "through line" and "cord circuit" use, and for large installations as well as for small ones. The signaling apparatus which may have different features in different

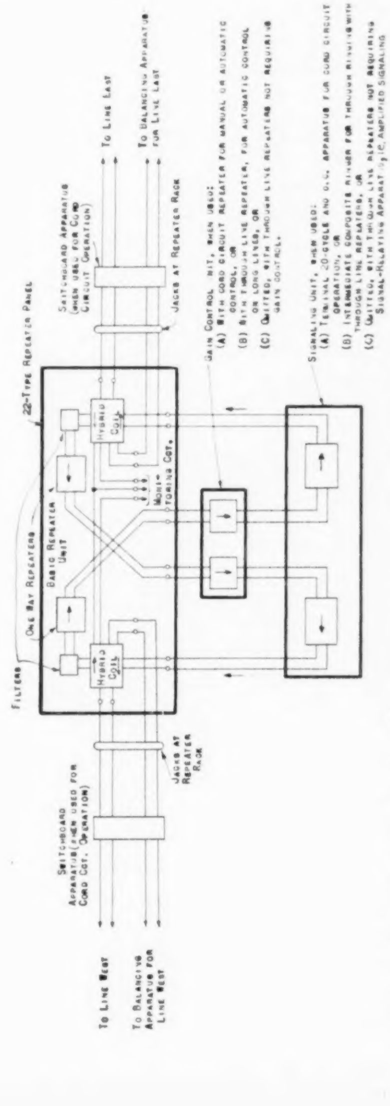


Fig. 8—Schematic Circuit Diagram Showing Two-Way Two-Element Repeater Arranged to Employ One Type of Basic Repeater Unit for all Classes of Service

types of offices, and which may not be needed with "through line" repeaters in some cases, will be furnished as a separate panel from the basic repeater unit. The apparatus which will permit the repeater to be used for cord circuit operation is also to be furnished as a separate unit and may be used in place of the through signaling unit, without changing the basic repeater. The filter which will determine the cut-off frequency of the repeater will also be furnished as a separate piece of apparatus mounted on the repeater set, thus the repeater may be suited to any desired type of line by providing the proper filter.

Another interesting phase of repeater development has been that in connection with the power supply for the vacuum tubes. This includes (1) a source of filament current (2) a source of plate potential and (3) a source of grid potential. In meeting the requirements of cable installations the principal improvements desired have included the use of batteries common to as many repeaters as possible, in place of individual batteries, closer regulation of potentials and the elimination of dry cells where practicable.

In some of the earliest installations a separate six-volt storage battery was used to supply the filament current for the vacuum tubes of each repeater set. Later, the filament current supply was taken from an 11-cell central office storage battery through a rheostat. As the potential of the 11-cell central office battery, normally 24 volts, varied from 20 to 28 volts during the operation of a charge and discharge routine, it was necessary to adjust the rheostat at frequent intervals to maintain constant current in the vacuum tube filaments. With the greatly increased number of repeaters per station which has occurred in cable systems, the maintenance involved in readjusting the filament currents would have become prohibitive on this basis. Accordingly, for the larger installations, duplicate 11-cell batteries normally floated from generators and provided with an emergency cell to maintain voltage during an emergency discharge are proposed. By this improved arrangement it is expected to be possible to maintain the filament voltage within one volt up or down from its normal value, even during an emergency discharge, until the batteries are almost completely discharged. This improved regulation will entirely eliminate adjustments of the individual repeaters during operation to secure proper values of filament current.

For the plate voltage supply dry cells have sometimes been used in small installations. These are now being displaced to a large extent, by small storage cells, two groups being used so that one group may be charged while the other is in service. In the large cable

installations, the current drain on the 130-volt plate batteries has sometimes reached values as great as four or five amperes, so that it is now planned to float these batteries also, instead of operating on the charge and discharge basis. It is expected by this means to obtain regulation of the plate voltage within plus or minus five volts from the normal value of 130, at all times.

Consideration is being given to another possible improvement in the power arrangements for large repeater installations. This involves the proposal to use a storage battery common to all of the repeaters for supplying the grid potential. If it is found that this arrangement is practical, it will permit the elimination of the individual dry cell batteries which have been employed, with a consequent saving

Fig. 9—Typical Storage Battery Room for Cable Repeater Station

in maintenance that might be appreciable. It seems likely that a very small storage battery would serve for this purpose since the current drain is negligible.

The amount of power required to operate the vacuum tubes in a large telephone repeater installation is considerable. Each filament requires a current in the neighborhood of one ampere and, while the filaments are connected in series in such a way as to utilize as efficiently as practicable the full potential of the central office battery, the load on this battery sometimes amounts to several hundred amperes.

Fig. 9 gives some idea as to the size of the storage batteries for a typical office of this kind. The large cells in wooden tanks are those making up the filament batteries, each of which is an 11-cell battery of 24-volt nominal rating. These two batteries in parallel are large enough to carry the office load for at least 12 hours in the event of a complete failure of charging equipment.

Fig. 10 shows the charging generators and power switchboard for a typical cable repeater office. The gas engine drives emergency generators to float the filament and plate batteries in the event that

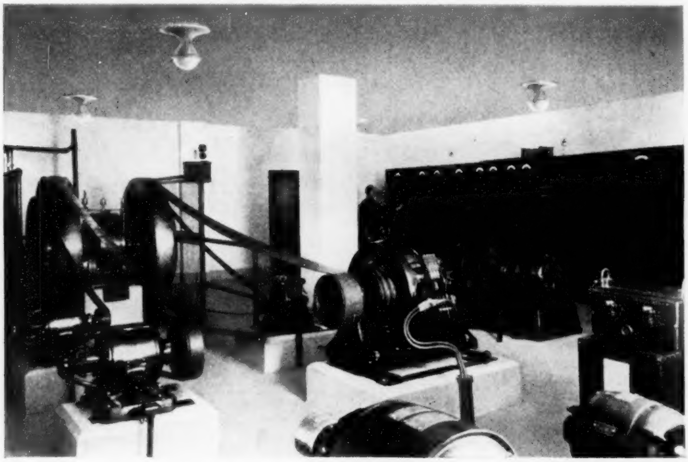


Fig. 10—Typical Power Room for Cable Repeater Station

the regular electric power supply service fails. Alarms are provided to indicate any abnormal condition such as a voltage higher or lower than normal, a blown fuse, etc. These are grouped in an annunicator cabinet on the wall.

Test Board Equipment

The test board forms an important part of the cable plant equipment, since it is the one point in the office where all of the lines and the equipment as a whole may be reached readily for purposes of testing or re-routing, in cases of line trouble or changes in layout. The form and arrangement of the test board equipment consequently have an important bearing on the effectiveness with which the facilities are handled by the maintenance forces. This is particularly true in the cable plant where the number of circuits involved is large.

In general, all toll line conductors are brought into a central office, from the outside, through a cable and first appear at a rack called a "distributing frame" where they are soldered to exposed terminal lugs so that they may be reached for the purpose of connecting them to apparatus within the office. This arrangement is well suited to the permanent connections but is not intended to permit frequent changes or the ready removal of the apparatus for line testing.

The necessity for rearranging the connections between the apparatus and the lines in cases of line trouble makes it desirable at times to be able to make such changes quickly, and the means which have been provided for this purpose are located at the "test board." Here both the line conductors and the apparatus units are wired to "jacks" which are arranged to permit the transfer of the normal connections by the insertion of plugs wired to flexible cords. A temporary connection made in this manner through a conducting cord wired to two plugs is called a "patch." The apparatus for determining the location of line trouble is also located at the test board and is wired to cords and plugs so that it may be connected to any line in the office readily, upon the occurrence of line trouble, without necessitating changes in soldered connections.

The test boards used for the open-wire plant have been designed to take care of 40 to 80 line conductors in one position, that is, in a board three feet long. The amount of testing and patching work required on open-wire lines has been such as to make this a convenient number of circuits to handle within this space. In cable installations, however, the amount of testing and patching per line conductor is less, while the number of circuits in such an office is much greater. Consequently, in cable offices it is possible to concentrate a larger number of wires within a given test board space. This is desirable from the standpoint of economy in space as well as from that of convenience in operation.

One of the first steps considered in the development of efficient test board equipment for cable installations has been the reduction of the number of jacks per circuit. In open-wire installations it has been the practice to equip each line circuit and each equipment unit, such as a composite set or phantom coil, with a full complement of jacks suited to provide the maximum degree of flexibility in "patching," thus permitting the ready interchanging of individual equipment units, lines and drop circuits. In cable installations, where the circuits are more likely to be uniformly equipped with the same types of associated apparatus and where the line troubles are less frequent, it is expected to be possible to eliminate certain of the jacks,

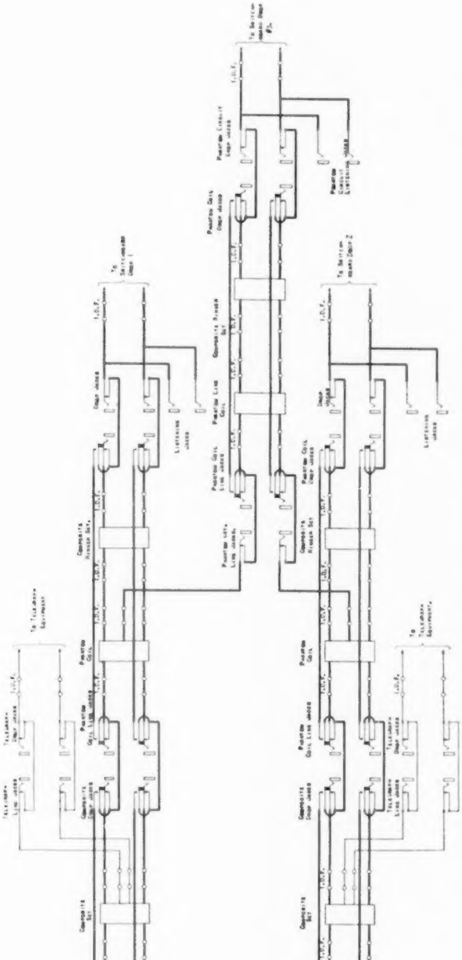


Fig. 11—Typical Phantom Group Circuit for Open-Wire Installations

such as those associated with the composite sets and phantom sets, thus simplifying the equipment for terminating the toll lines. Fig. 11 shows the typical open-wire arrangement for a terminating phantom group circuit in which the maximum number of jacks is furnished. This requires a total of 46 jacks. Fig. 12 shows the arrangement of a terminating phantom group circuit as planned for a cable installation. In this case a total of but 30 jacks is required.

Another important development expected in the test board arrangements to suit them to cable use is the grouping together of the jacks serving similar functions. Considerable improvement in

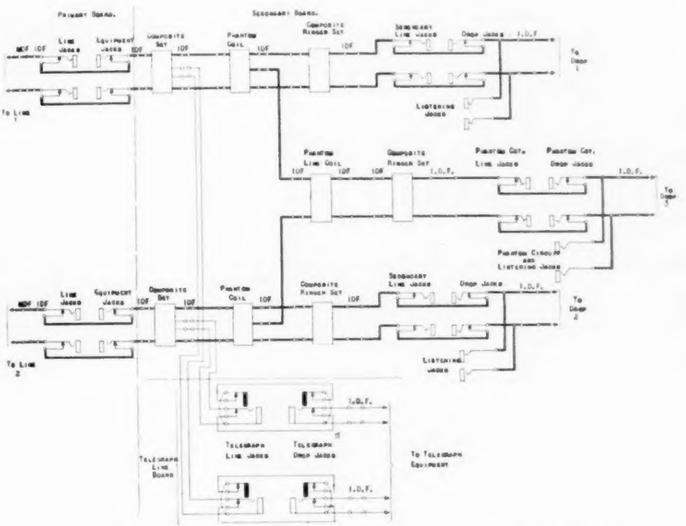


Fig. 12—Typical Phantom Group Circuit for Cable Installations

operation is thought to be possible with the jacks having different functions located at different test board positions. In this way all of the line conductor jacks, for example, may be assembled together in consecutive order, and since several hundred of these may be involved in a single installation, this should greatly facilitate the identification of the desired circuits by the attendant in the process of patching and testing. This grouping of the jacks will also effect a saving in testing equipment, since it will eliminate the need of the line testing apparatus, such as the Wheatstone bridge, at positions where the line conductors will not appear.

Fig. 13 illustrates both the open-wire and the proposed cable methods of grouping the jacks. In the arrangement for open-wire circuits the jacks associated with both the lines and equipment are located adjacent to each other in the same test board panel. In cable installations the jacks having similar functions are to be grouped

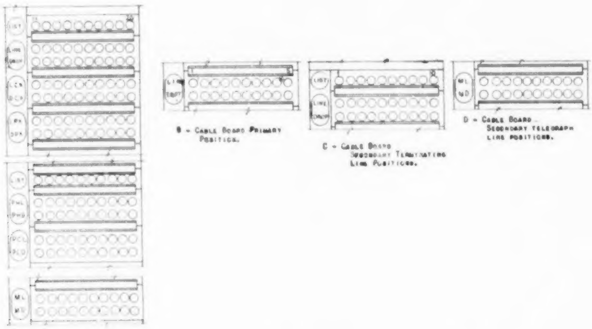


Fig. 13—Typical Jack Assembly Arrangements at Test Board

together, the groups having different functions being mounted in different panels which may be located in different test board positions. These various groups, in the latter case, are planned as follows:

1. Primary line testing position for testing and "patching" toll lines only. This position to be used for locating faults in the cable circuits and equipped with a Wheatstone bridge and voltmeter, to permit the necessary electrical measurements for this purpose. This position is also to be used for making temporary changes in the assignments between the lines, and the equipment as a whole, but is not to be arranged to permit changes in individual equipment units, such as composite sets, phantom sets, etc. The jacks to be located at this board are to include those designated as "line jacks" and "equipment jacks," in Fig. 12.

2. Secondary terminating line positions for testing and "patching" the lines between the "drop" side of the equipment and the toll switchboard circuit. This position is to be used for determining the general nature of a trouble and its general location, i.e., whether it is in the direction of the line or in the direction of the "drop," and for clearing troubles not requiring line tests. This position is not to be equipped with Wheatstone bridge testing

apparatus as at the primary board. The jacks to be located at this position are to include those designated as "secondary line jacks," "drop jacks" and "listening jacks," in Fig. 12.

3. Secondary telegraph line positions for testing and "patching" the telegraph line circuits. This position is to be used solely for interchanging telegraph lines and telegraph equipment, in cases of temporary changes in assignment and is not to be equipped with the line testing apparatus. This position will permit changes to be made in the telegraph assignments without interfering with the telephone circuits. The jacks to be located at this position are to include those designated in Fig. 12 as "telegraph line jacks" and "telegraph drop jacks."

A further and more extensive improvement in test board design is anticipated as a result of development work whereby it will be possible to employ panel mounted keyshelf equipment units, jacks and testing apparatus, which in the standard board are now housed in a

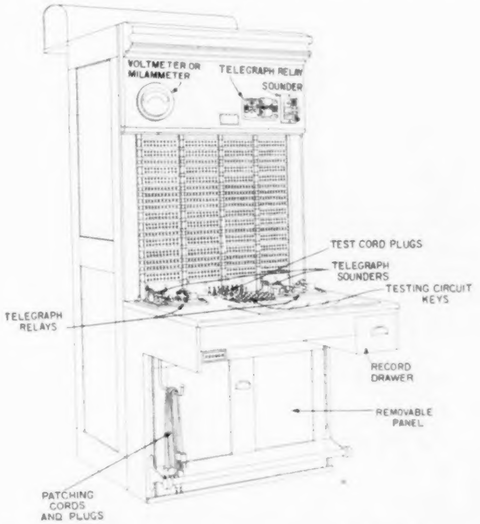


Fig. 14—Typical Assembly of Floor-Mounted Test Board—One Position

large wooden section. This will have the advantage of uniformity with the other toll equipment, as well as requiring less space. It will also permit flexibility in the use and installation of the various combinations of keyshelf equipment, jack equipment and other testing

apparatus which may be required to suit each case. While this arrangement will have its chief advantages when applied to large cable installations, it will also be well suited to open-wire use and small installations, since its design will permit the highest degree of flexibility with respect to both the amount and type of equipment.

These points may be illustrated by comparing the general features of the two types of boards. Fig. 14 shows the assembly of a one-position section of the present type of board employing a wooden

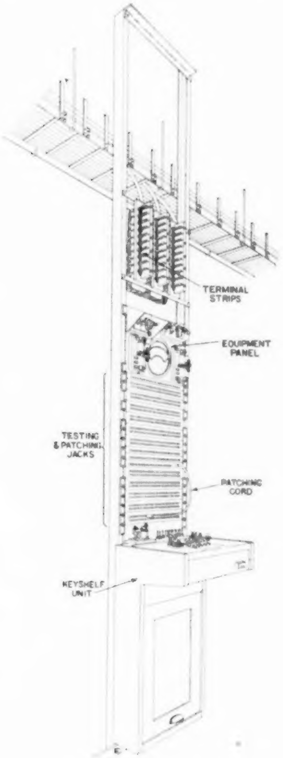


Fig. 15—Typical General Assembly of Panel Mounted Test Board

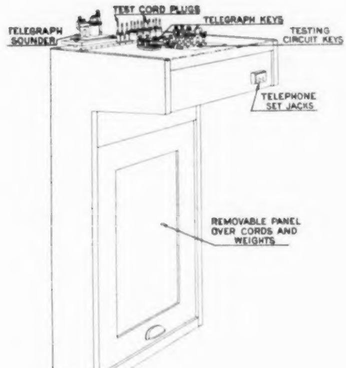


Fig. 16—Typical Assembly for Keyshelf Equipment Unit for Panel Mounted Test Board

framework. This board, with the necessary allowances for aisle space, requires a floor area amounting to about 24 square feet, while it houses a maximum of about 1000 jacks corresponding roughly to about 40 jacks per square foot. Fig. 15 shows a typical position em-

ploying the proposed panel mounting method. Such a board will occupy a floor space of about 10 square feet and will take care of about 600 jacks, corresponding roughly to a capacity of 60 jacks per square foot.

This latter type of board is to be made up of a number of panel units which are to be assembled on two vertical supports. The principal types of units to be provided for the purpose are the keyshelf units, the jack mountings and the equipment panels which may be combined together as desired to give the necessary facilities.

Fig. 16 shows a typical keyshelf unit designed for the panel type board. By constructing the keyshelf unit as a separate piece of apparatus, it is expected to be possible to standardize the necessary types of keyshelves to fit all ordinary field conditions and to specify the desired type of keyshelf to go with any particular arrangement or number of jacks. The number of keyshelf units of any given type may be as desired for each installation, thus the proportion between the jacks and the keyshelf equipment may be suited to each type of office.

Fig. 17 shows a typical arrangement of the jack equipment and the mountings which are to be employed for the jacks. This type of jack mounting will make it possible to mount the jacks on the same sup-

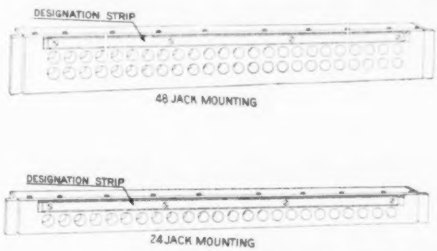


Fig. 17—Typical Assembly of Jack Mounts for Panel Mounted Test Board

ports as the testing equipment. The mountings are to be attached to the supports by fasteners, each occupying a vertical space of $1\frac{3}{4}$ inches and drilled to fit the usual drillings in the supports. This will permit the close association of the jacks with the desired testing apparatus. It will also be possible, by this means, to use for the jacks only such of the available vertical space as may be desired, the remainder being used for other equipment, thereby effecting economy in the use of the space. The arrangement is thus expected to be advantageous both in large installations, where the various groups of

jacks are desired to be arranged at separate primary and secondary positions, and in small installations in which but a few jacks may be required for all purposes.

Fig. 18 shows the proposed assembly of a typical panel equipped with voltmeter, telegraph relay and sounder, such as are usually mounted above the jack field. The advantage of using the panel

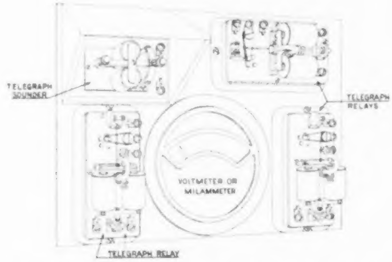


Fig. 18—Typical Assembly of Apparatus Panel for Panel Mounted Test Board

mounting for this equipment is expected to be the same from the standpoint of convenience and economy as that obtained from this method of mounting the keyshelves in relation to the various jack fields. It will permit the location of this equipment at any desired position and eliminate duplication at positions where this apparatus is not necessary. It will also permit flexibility in regard to the type of panel associated with a given jack field.

Signaling Equipment

The principal purposes of the signaling equipment in telephone lines are (1) to permit a subscriber to signal the central office operator, as he does automatically in removing the receiver from the switchhook, (2) to permit the central office operator to ring the subscriber's bell and (3) to permit the operators at different central offices to signal each other. At repeater stations this equipment serves to pass the signals around the telephone repeaters which might otherwise interfere with their transmission.

The switchboards in both the local and toll central offices have been provided with a source of 20-cycle current for signaling, as current of this frequency is suited to operate the subscriber's bell directly. This frequency has also been satisfactory for operating the signaling apparatus in the various local trunk circuits and in short toll lines without superposed telegraph. The introduction of superposed telegraph and telephone repeaters, however, has prevented

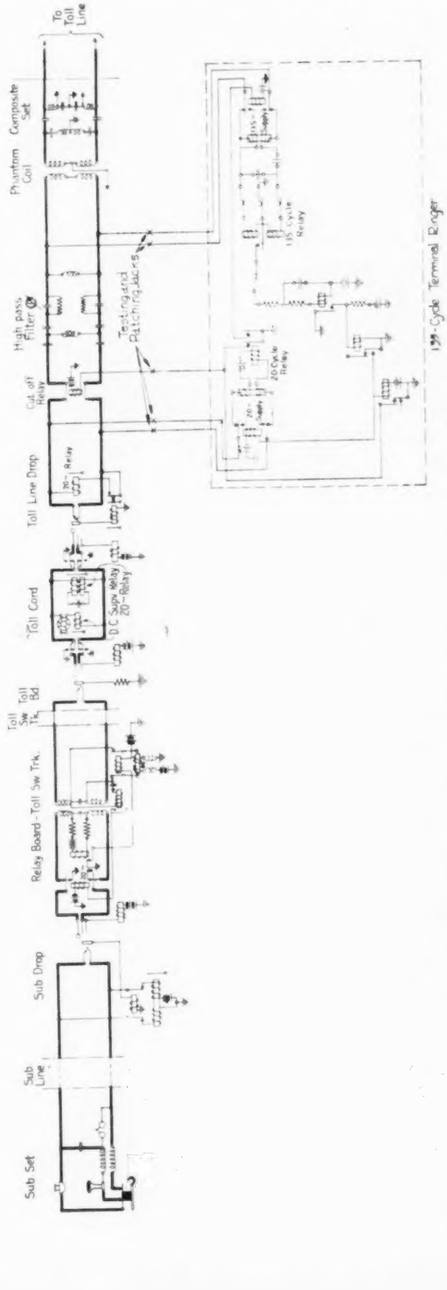


Fig. 10—135 Cycle Ringdown Toll Signaling System—Simplified Diagram

the use of 20-cycle signaling on many toll circuits. It has accordingly been expedient to use a higher frequency for the majority of these circuits, and the one which has been most generally used for this purpose has been 135 cycles. It is desired, therefore, to bring out here some of the more important features of the 135-cycle signaling arrangements which are being developed for cable systems.

The signaling equipment for cable systems has been required to meet more severe conditions than those ordinarily encountered on open-wire lines. In order not to interfere with the direct current telegraph system used on the cables, it has been necessary to limit the signaling current to a few milliamperes. Furthermore, the characteristics of the cable apparatus have been such as to attenuate the signaling currents to a greater extent than in open-wire systems. It has accordingly been necessary to undertake the design of signaling equipment of greater sensitivity, as well as to provide a source of supply of signaling current possessing a high degree of freedom from harmonics and capable of being closely regulated.

The desired increase in sensitivity of the signaling system is expected to be obtained through the use of a highly selective circuit in conjunction with a very sensitive 135-cycle relay. Fig. 19 shows the general scheme of the circuit which is being developed for the purpose.

It is seen that this includes a filter which may be inserted in the line circuit between the signal-receiving apparatus and the switchboard "drop" to give the desired terminal impedance at the signaling frequency, thus preventing the low impedance shunt on the signaling relay which would otherwise be caused by the "drop" circuit at a frequency as low as 135 cycles. The filter is arranged so that it need be inserted in the circuit only in such cases as may require the increased signaling range obtained in this manner, thus, it may be omitted on the shorter circuits where it is usually unnecessary.

The general circuit arrangement shown in Fig. 19 will permit the ready interchange of different types of signaling apparatus. With this arrangement of the apparatus, any "ringer" of any desired frequency combination such as 20–135 cycles, 20–20 cycles, 135–135 cycles, etc., may be connected temporarily to the system when desired, by means of "patching" cords, without requiring any changes in the permanent wiring.

Much of the sensitivity and selectivity which may be obtained with this signaling system are due to the design of the 135-cycle relay. As shown in Fig. 20, the relay is designed to make it capable of close

and accurate adjustment. Both the magnetic air gap and the contact spacing may be adjusted, about 0.0015 inch or 0.038 millimeter being used ordinarily for the latter. The relay is mechanically tuned, the natural period of the reed with its adjustable weight corresponding closely to the frequency of the signaling current. A stop

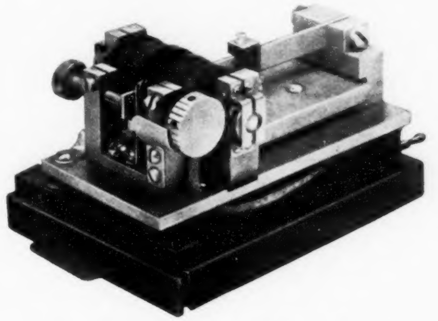


Fig. 20—Assembly of 135-Cycle Relay

pin is provided which prevents undue vibration of the reed due to transient impulses or excessive currents. Also, the relay circuit is electrically tuned by a shunt capacity and a series inductance and capacity. It is thus very selective and is relatively free from the ordinary sources of interference such as those caused by mechanical vibration, telegraph signals, switchhook impulses, voice currents, etc. The sensitivity of the relay is such that it will operate on as little power as 30 microwatts, corresponding to a current in the neighborhood of 0.25 milliampere.

This relay is to be mounted so that it may be inserted in the circuit or removed from it in the manner of a plug and jack, without requiring changes in the permanent wiring and without affecting the circuit operation, except to interrupt the signal-receiving system, while the relay is removed. Thus it will be very convenient to make the necessary adjustments of the relay separate from the circuit with which it may be associated in service. The relay will be well protected from mechanical interference, such as building vibration, by padding in the mounting which prevents rigid mechanical connection between the relay and its external support.

The assembly arrangements proposed for the new signaling apparatus are such as will fit in closely with the panel mounting methods designed for the remainder of the toll equipment in cable in-

stallations. Fig. 21 shows the assembly proposed for the new composite ringer set. This method of mounting the composite ringer

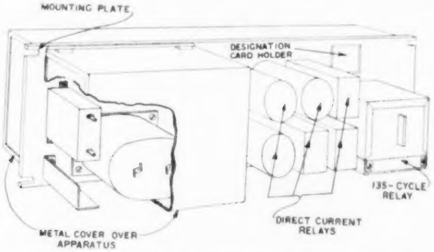


Fig. 21—Typical Assembly of 135-Cycle Composite Ringer Set

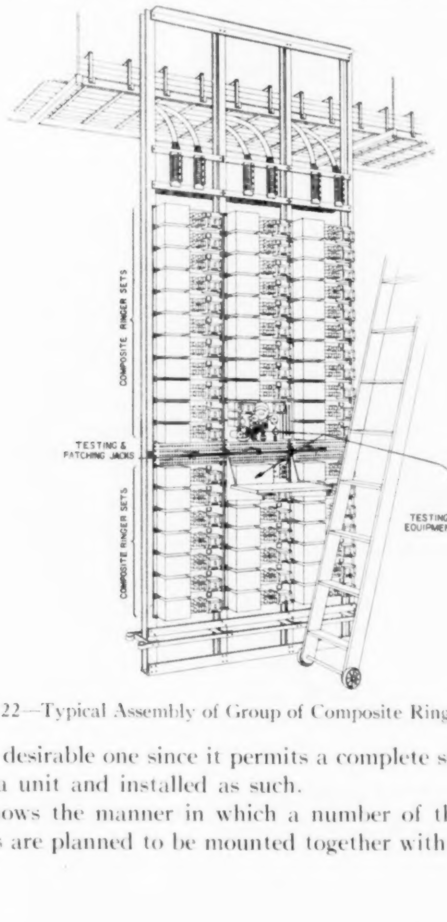


Fig. 22—Typical Assembly of Group of Composite Ringer Sets

set is a very desirable one since it permits a complete set to be manufactured as a unit and installed as such.

Fig. 22 shows the manner in which a number of these composite ringer panels are planned to be mounted together with the associated

testing equipment in a large installation. This close association of the composite ringers with the jacks and testing equipment is expected to be of considerable advantage in facilitating the maintenance of the apparatus.

The need in cable systems of a particularly well-regulated source of 135-cycle signaling current, with sufficient capacity for a large number of lines, has made it desirable to undertake the development of both a special type of interrupter and a motor-generator for this purpose. Close frequency regulation is also very desirable in signaling over cable circuits, in view of the increased sensitivity in receiving which may be obtained by the use of very selective receiving apparatus.

Fig. 23 shows the circuit arrangement of the interrupter which is expected to be provided for this purpose. This includes a vibrating

Fig. 23—135-Cycle Interrupter Circuit—Simplified Diagram

reed actuated by an electromagnet when direct current is applied. The contacts on the reed being in series with the battery circuit, intermittent operation is secured in the manner of an ordinary buzzer, the speed of operation for a given applied voltage being de-

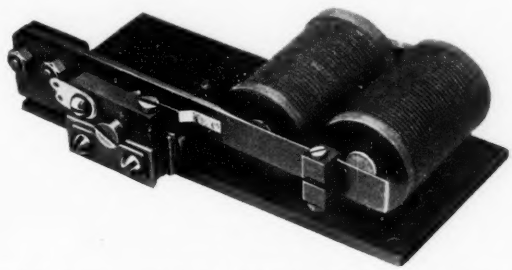


Fig. 24—Assembly of Vibrator for 135-Cycle Interrupter

termined by the natural period of the reed. The actuating circuit of the vibrator also includes the primary side of a transformer, the secondary side of which is connected to a filter for suppressing harmonics in the output. The maximum output capacity of the interrupter is about three-fourths of a watt.

Fig. 24 shows the vibrating element to be employed in this interrupter. This is equipped with an adjustable weight so that the frequency of the output may be regulated. For an input voltage variation of 20 to 28 volts, the output frequency will vary about 5 cycles.

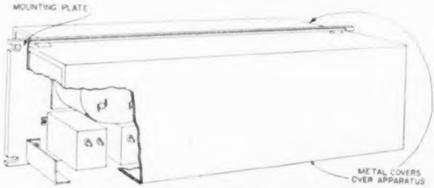


Fig. 25—Assembly of 135-Cycle Interrupter

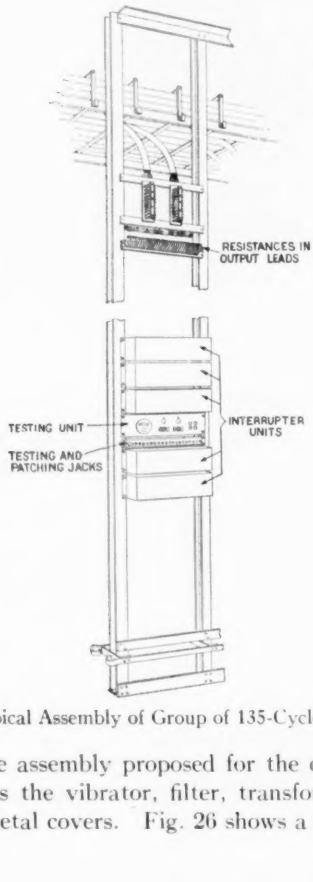


Fig. 26—Typical Assembly of Group of 135-Cycle Interrupters

Fig. 25 shows the assembly proposed for the complete interrupter unit. This includes the vibrator, filter, transformer, etc., mounted on a panel under metal covers. Fig. 26 shows a typical arrangement

of a group of interrupters with the associated testing equipment. These assembly arrangements are uniform with those previously described for other types of equipment.

The 135-cycle motor-generator developed for signaling purposes is shown in Fig. 27. This outfit includes two motor-generators, one for regular service and one for reserve, on one panel, the control

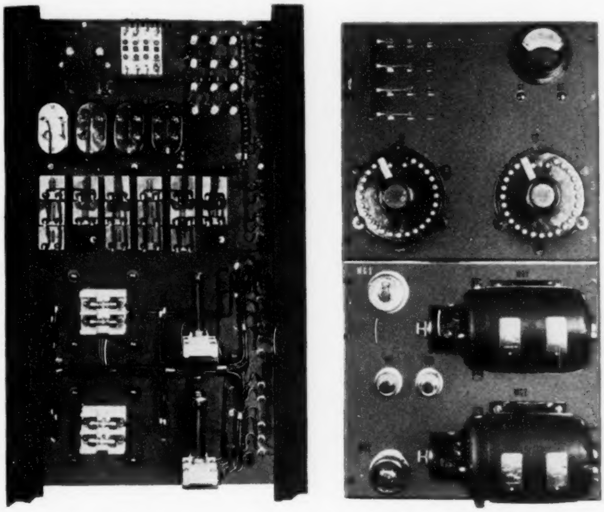


Fig. 27—Assembly of 10-Watt, 135-Cycle Motor-Generator Set and Associated Apparatus

equipment being mounted on a second panel. The output of the motor-generator, which is provided with a filter, is practically free from harmonics. Its voltage range is from 30 to 40, while its frequency range is from 133 to 137 cycles, including full load to no load conditions. The output capacity of this machine is approximately 10 watts.

General Assembly Arrangements

In the preceding descriptions some mention has been made of the panel assembly method which is expected to be used with much of the cable equipment. It might be well to speak briefly here of some of the features of this mounting method which are expected to have general application.

The large number of equipment units per station in the cable plant has been one of the principal factors in determining the requirements

for efficiency in the design of the equipment. These requirements have been mainly (1) compactness in dimensions (2) uniformity in assembly arrangements and (3) simplicity in design.

To make possible the housing of the equipment for a cable installation within a building of reasonable dimensions, it has been necessary to economize carefully in space. Fig. 28 shows in a general

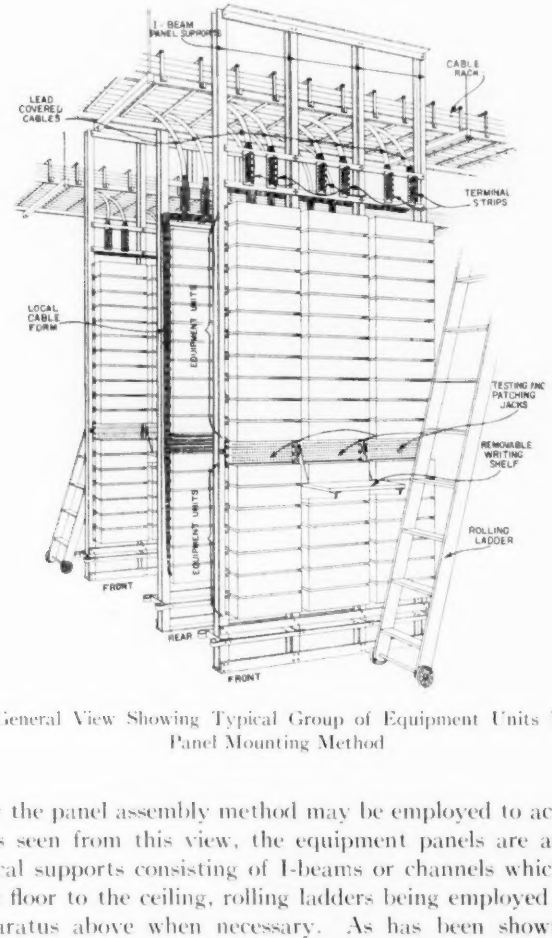


Fig. 28—General View Showing Typical Group of Equipment Units Employing Panel Mounting Method

way how the panel assembly method may be employed to accomplish this. As seen from this view, the equipment panels are assembled on vertical supports consisting of I-beams or channels which extend from the floor to the ceiling, rolling ladders being employed to reach the apparatus above when necessary. As has been shown in the preceding views of individual equipment units, all of the adjustable apparatus is mounted on the front of each equipment panel and the

material not requiring adjustment when in operation is mounted on the rear, since by this means the rear aisles can be made narrower than the front aisles. Thus, all available space is utilized to the extent that the floor area required per equipment unit in a cable installation will be as small as the maintenance, manufacturing and installing requirements for the apparatus will permit.

In view of the many different types of equipment units having a variety of special functions which are required to make up a complete cable installation, a high degree of uniformity in design is necessary to permit efficiency in their use. It is expected that this will be accomplished effectively by applying the panel assembly method in a uniform manner to practically all types of equipment units. All panels are to be of a uniform length, designed to mount on vertical supports spaced $19\frac{1}{2}$ inches between centers. The height of the different panels will vary, according to the amount of apparatus in each unit, but this vertical dimension is in all cases to be a whole multiple of $1\frac{3}{4}$ inches. By applying these specifications widely, it will be possible to secure interchangeability between panels and to employ uniform methods in grouping the different units, thus facilitating their installation and use.

The simplicity of the design of the equipment units comprising the various panels is also of great importance. This has necessitated careful attention to the forms of the individual pieces of apparatus, in order that these might fulfill their specific functions efficiently while at the same time fitting in well with the general equipment arrangements. To this end, new types of apparatus, such as repeating coils, retardation coils, etc., are being developed especially for cable use. Much will also be accomplished toward the simplification of the panels by carefully avoiding duplication in the accessories to the different types of equipment and by dissociating from the individual units of all types any pieces of apparatus capable of being made common to a number of units, or subject to different methods of application in different types of offices.

Other important advantages are anticipated in the panel assembly method. One of these is that it will permit the assembling together on one panel of all of the pieces of apparatus of different types which may be desired to form a distinct equipment unit. In large installations, completely equipped racks including a number of equipment units with the associated testing apparatus may be assembled in the factory on the supporting uprights and wired to the terminals at the top of the rack, thus simplifying the installation work. The location of the testing apparatus on the same rack with the equipment

panels has the further advantage that it will place the apparatus to be adjusted within easy reach of the testing facilities. Furthermore, the testing apparatus will serve for the maintenance of a greater number of equipment units when all of the space between the floor and ceiling is utilized, thus reducing the amount of testing equipment required.

Radio Extension of the Telephone System to Ships at Sea[1]

By H. W. NICHOLS and LLOYD ESPENSCHIED

SYNOPSIS: The paper describes the development of a two-way radio-telephone system and its use in extending the Bell Telephone System to connect with ships at sea. The electrical considerations and the experimental work involved in determining the system-design of the radio link are discussed. Two land stations were established, one of them a permanent three-channel station on the New Jersey coast. Two coastal vessels and finally one trans-Atlantic liner were equipped. These installations are briefly described in the paper.

The operation of the combined radio and wire system is explained, particularly in respect to the transmission characteristics of the over-all system and the effect thereupon of the movement of the vessel and of variations in atmospheric conditions. Measurements of the variations in the field strength received from field vessels at sea show why it is possible to receive over very long distances at favorable times at night and not during the day. The method of establishing combined radio-telephone-wire circuits to ships is described and representative results are given of the considerable telephone traffic which was handled over the system experimentally during a period of trial operation. Tests of multi-channel telephone operation to several ships through the Deal Beach shore station, and also tests of simultaneous telegraph and telephone operation from the same vessel are described. Connection of a vessel thru the transcontinental telephone line to the Catalina Island radio-telephone system, whereby the vessel in the Atlantic talked with an island in the Pacific, is briefly described, and finally the outstanding conclusions of the entire development work are given.

IN 1919, the American Telephone and Telegraph Company and the Western Electric Company initiated a development program which had for its object the development of a radio telephone system capable of enabling the service of the Bell Telephone System to be extended to include vessels at sea. The program involved extensive development work in the laboratory and field, the establishment of shore and ship stations, and the putting of the system into practical operation, altho on a limited and experimental scale.

It is the purpose of this paper to describe the results of this development work from the standpoint of the complete system, with emphasis upon the general transmission and operating features rather than upon the details of the apparatus developed to perform the necessary functions. The development divides itself, naturally, into two parts: first, the determination of the system-design and the establishment of the necessary stations, and, second, the study of the transmission and operating characteristics of the system.

[1] Presented before The Institute of Radio Engineers, New York, January 3, 1923. Reprinted with minor changes from the Procd. of Institute of Radio Engineers, June, 1923.

Part I

RADIO SYSTEM-DESIGN AND ESTABLISHMENT OF STATIONS

General Plans

The fundamental condition laid down at the beginning of this work was the very general one that there should be developed a system by which any telephone subscriber of the Bell System could carry on a conversation with a telephone station located on a ship, and that, from the point of view of the speakers, the operation should be similar to the carrying on of an ordinary toll call between land wire subscribers. This, of course, involves the development of a satisfactory two-way radio telephone system for ship use. Furthermore, it was desired to be able to carry on three simultaneous and independent conversations between three ships and one land station, since a final commercial system will involve the establishment of several circuits simultaneously. These 2-way transmissions were to be obtained without employing an excessively large frequency band.

A rough study of the problem resulted in a decision to locate the experimental land stations about 200 or 250 miles (320 or 400 km.) apart and to try for reliable commercial transmission to ships at a distance of approximately 200 miles (320 km.).

The transmission problems involved in this work, which were different from those in wire telephone engineering, were:

(a) A much greater variability in the transmission equivalent to be expected in the radio link;
(b) A much greater and more variable interference, both natural and artificial;
(c) A lack of secrecy in the sense of a wire system;
(d) Greater possibilities of cross-talk between channels because of the use of a single medium;
(e) More complication in the matter of signaling and in the setting up of the telephone circuit.

The apparatus problems were, of course, entirely different from those of wire transmission and will not be considered in detail in this paper.

An engineering project of this kind divides itself naturally into two phases; that of the development in the laboratory of systems and apparatus which are technically suitable for the work and, second, the providing in the field of a model system, incorporating the knowl-

edge obtained in the laboratory as a means for enabling the system to be tried out. A preliminary survey of the purely technical problems convinced us that the more important ones were the development of two-way radio telephone apparatus and of multi-channel systems which would operate from a single transmitting station without interference between channels; the design of transmitting apparatus which would satisfy the requirements and which could be built with the vacuum tubes available and the development of a type of receiving system which would provide sufficient selectivity to allow an economical use of the frequency range and at the same time fit in with the two-way system most likely to be adopted. It was decided that during the laboratory development work preparations should be made in the field for providing the necessary experimental stations. This field work as it developed included the location of station sites, the actual construction of the station buildings and the antennas, the equipping of the stations with the apparatus as developed in the laboratory and as further developed in the station, the equipment of the ships, and, finally, the operation and tests of the overall system.

System Design Considerations

In the beginning it was thought that to cover the required 200 miles (320 km.) range about one or one and one-half kilowatts in the antenna would be necessary. It was not known that wave lengths would be made available for this work by the Department of Commerce. To produce this amount of power in the antenna there were available Western Electric 250 watt tubes which it was decided to employ. The question then arose as to the particular type of transmission systems most suitable for the work. The points of importance in solving this problem are as follows:

The greatest economy both in power and in wave length range may be secured by transmitting only one side band of the modulated wave. Moreover, this method has the great advantage that variations in the transmission characteristics of the medium do not cause as great fluctuations in the received signal. This is because the received signal is proportional to the product of the carrier and side bands and if the carrier is supplied locally instead of being transmitted, it is not affected by transmission factors. The use of such a system, however, or of one in which only the carrier is suppressed, throw upon the receiving set the burden of maintaining a constant oscillator frequency not only complicating it but also making reception impossible for the great majority of ships which are equipped with only straight detectors. This would defeat general inter-communica-

tion in emergency. Further, it practically restricts the transmitting set to one in which the power tubes are used as amplifiers, and it was known that some difficulty might be experienced in operating a number of 250 watt tubes in parallel if it should be necessary to transmit at wave lengths as low as 300 meters. For these reasons and after some development work it was decided that the proper system to use in the first experiment was one in which modulation is carried on by the constant current method which requires about an equal number of modulator tubes and power tubes and sends out all components of the modulated wave.

The simultaneous transmission of three channels from the land station may be accomplished in several ways. It is possible, for example, to carry on such multi-channel operation from one antenna, which is multi-tuned, or from three separate antennas. The antenna power may be supplied by one system of tubes carrying all three conversations or the power tube system may be split into three parts. Also, using a single antenna simply tuned it would be possible to transmit the three channels from one system of tubes by a system of double modulation which had been installed by the Western Electric Company on United States battleships two or three years earlier. The difficulties which are likely to arise in these various schemes are as follows: The use of multi-tuned antennas involves loss of power in the circuits used to give the antenna three degress of freedom. The use of a single system of power tunes for three channels requires that the tube system be capable of handling a large overload at times without impairment of quality, since it is possible that the peaks of three channels may occur simultaneously. It was expected that under conditions of this kind there would be inter-modulation of the channels due to the modulating action in the plate circuits of the power tubes. The use of three separate antennas located very close to one another and tuned to frequencies differing by three or four per cent. might lead to such close coupling of the three channels that cross-talk and modulation of one channel by another would result, the latter by plate modulation of one set of tubes by the currents induced in its antenna. The use of the double modulation system—altho requiring but one radiated carrier—is open to the objection of overloading and cross modulating of channels and also to the objection that the receiving apparatus aboard ship must be more complicated. An analysis of these and other proposed methods of operation resulted in the decision to employ at that time three separate but closely adjacent antennas and three separate transmitting sets using the constant current modulation system. This choice was made because of conditions peculiar to this particular problem and to the vacuum

tubes then available. Of course, improvements can be made in the system at the present time as a result of the information obtained in the development using the very much larger vacuum tubes now available.

These decisions, therefore, determined the general type of system to be used, namely, one in which many of the known advantages of single side band transmission were sacrificed in order to secure simple apparatus, to make use of then existing power tubes and to enable the transmission to be received generally.

The problem of securing the two-way operation necessary aboard ship and for combined radio and wire operation may be attacked in several ways. In general, there are three methods available:

(1) In which the east and west channels are established alternately and not simultaneously, by switching. The push-button scheme is a familiar example, although unsuitable for tying in with the wire telephone system. Another arrangement is the use of voice-operated relays to throw the terminal apparatus into the sending or reciving condition, depending upon the direction of transmission.

(2) The use of the principle of balance to separate the outgoing from the received transmission. The radio receiving antenna circuit is balanced with respect to the transmitting antenna circuit.

(3) Employment of different frequencies for the two directions of the two-way transmission, relying upon frequency-selecting circuits for affecting separation. The first two methods allow of operation on the same or on different carrier frequencies.

All of these fundamental methods were considered in their several possible embodiments, and compared from the standpoint of the conditions to be met in the radio system itself and in linking it with a public service telephone system. The system finally adopted employed different frequencies for sending and receiving and secured discrimination by frequency selection supplemented at the land station by a moderate degree of special separation and balance. By using sharply selective receiving circuits, a moderate frequency difference between east and west channels sufficed to give the necessary degree of separation.

Preliminary Tests

By the time these decisions had been made there was available for experimental purposes a plot of land near Cliffwood, New Jersey. It

was decided to construct a model of the proposed antenna system on this plot and to operate small transmitting sets to determine the cross-talk and other important conditions. The antenna system decided upon consisted of three poles arranged in the form of an equilateral triangle supporting three antennas—one from the middle of each span to the transmitting shack at the center of the triangle. The dimensions of this model system were 50 meters (164 ft.) by 10 meters (33 ft.) high. Three experimental transmitting sets of small power were set up under the antenna system and studies were made of the interference produced between channels when all three channels were in use. Three receiving sets of the general form proposed were built and taken to a location near Elberon, New Jersey, about 16 miles (26 km.) from Cliffwood, at which place it was decided to locate the three-channel receiving station to co-operate with the New Jersey transmitting station a mile (1.6 km.) away.

In November, 1919, the first test of a three-channel system was held between Cliffwood and Elberon with the result that the receiving sets resolved conversations on carriers of frequencies of 725, 750, and 775 kilocycles without any cross-talk altho the received volume was so large as to be audible all over the room. This is a frequency difference between channels of approximately three per cent. A change in frequency to 747, 759, and 777 kilocycles resulted in a barely perceptible cross-talk on the middle channel, with no cross-talk on the others. These results indicated that the loop receivers which had been developed were sufficiently sensitive and selective to carry out the proposed three-channel work; and, altho a great deal of development work was done later on the receiving sets, the general principles were retained. It was found that some reliance must be placed upon the directional properties of the loop antennas, and considerable care was used to secure very sharp directional selectivity. This was done by compensating for the vertical antenna effect of the loop by a balanced connection to ground.

During the whole course of this ship-to-shore work very little trouble was experienced thru interference by continuous wave stations, even when their frequencies came within two or three per cent. of those to be received. We did, however, have much difficulty due to interference from spark stations, since they inherently occupy a wide frequency range.

Provision of Stations and Development of Apparatus

During the time the model system was being constructed at Cliffwood, land had been purchased at West Deal, Monmouth County,

New Jersey, for the permanent transmitting station. This station as it now appears is shown in Fig. 1. The permanent building was preceded by a temporary structure to house an experimental transmitting station which could be used as a model for the design of the four final sets to be located in the permanent building.

Preliminary studies were also made to determine the proper form to give to antennas suitable for three channel operation without excessive cross-talk, and this study indicated that by the use of series inductance and capacity the antennas could be stiffened enough to prevent excessive coupling effects and still pass the required frequency band.

While this work was going on, a two-way telephone set for use aboard ships was developed, and in the spring of 1920 one of these

Fig. 1

sets was installed aboard the steamship *Ontario* of the Merchants and Miners Transportation Company. Experimental communication with this ship, by means of the model transmitters at both Deal Beach and Green Harbor stations, showed that commercial operation, at least for one channel, could be maintained.

By the fall of 1920, the construction work on the four transmitting and receiving channels was completed and early in December a

demonstration of simultaneous three-channel operation from this station to ships was carried out with satisfactory results.

Fig. 1 shows a general view of the outside of the transmitting station at Deal Beach. The three steel towers form an equilateral triangle of sides five hundred feet (150 m.) and each is one hundred and sixty-five feet (50 m.) high. Steel cables to support the antennas are strung between these towers and also three cables extending inward support a fourth antenna which rises directly from the building in the middle of the triangle. One antenna goes to the middle of each of the first mentioned steel cables, so that there are a total of four transmitting antennas. One of these is intended for use at six hundred meters. The building is thirty by ninety feet (9.1 by 27.3 m.) and two stories high. The southern half comprises the operating room which rises two full stories. The other part of the building is taken up by an office, shop, power room, living and dining room and kitchen, and by six bedrooms.

Description of the Experimental Radio Stations

The system as developed at Deal Beach consists of four transmitting sets, operating into four separate, altho naturally coupled, antennas, one set and antenna being intended primarily for 600 meter calling and for emergency. The receiving station co-operating with Deal Beach is located about a mile (1.6 km.) north of that station and contains four receiving sets receiving energy from four loop antennas. The transmitting sets are capable of putting about one kilowatt of modulated radio frequency power into each antenna and are controlled from a telephone switchboard into which run trunk lines from New York City. A ten-pair telephone cable connects Deal Beach and Elberon and another telephone switchboard at Elberon permits the transfer of received signals back to the wire line. The radio station operates, therefore, generally as a telephone repeater arranged for two-way operation with two repeaters. At the ship stations, because of the small amount of space involved, transmitting and receiving was accomplished on the same antenna at different frequencies in the two directions. Because of the better receiving conditions on the shore the proper transmission balance was obtained by making the output of the ship transmitting set about one-quarter that of the land station.

The general principle of operation of one channel of the wire-to-radio repeater will be described from the schematic circuit diagram of Fig. 2, which shows, in the dotted blocks, one channel of the transmitting station, a ship station, and one receiving set. At the transmitter station the master oscillator, very carefully shielded to main-

tain constant frequency, operates into a two-stage amplifier, the last stage being fifty watt tubes, and from there into a bank of six radio frequency power tubes, each with a rating of 250 watts plate dissipation. Speech to modulate this radio frequency output enters from a telephone line and is applied to a speech amplifier the output of which operates into a bank of 250 watt modulator tubes in parallel. Thus both the radio frequency and the speech frequency currents are brought up to the high power level before modulation takes place. The six radio frequency and six speech frequency tubes have their

TWO-WAY RADIO-WIRE SYSTEM

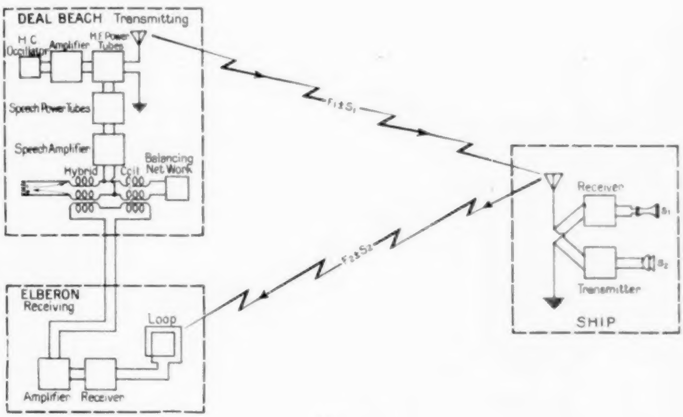


Fig. 2

plate circuits connected together and operate as a constant current modulation system. Thus current of the frequency F_1, generated by the master oscillator and amplified and modulated, is radiated from the antenna. The notation $F_1 \pm S_1$ indicates the radiation of the carrier and two side bands from this antenna. The incoming speech S_1, as it comes from the telephone line, passes through the hybrid coil and to the balancing network shown. This balancing network has an impedance characteristic similar to that of the incoming line, and the combination of hybrid coil and network is similar to that used in telephone repeater practice to secure two-way operation. The object of this arrangement is, of course, to prevent signals, coming in from the receiving station, operating upon the transmitter of the outgoing channel. If the balancing network is an exact picture of the incoming line and if the hybrid coil is properly made, incoming

signals for transmission west on the telephone line will produce no voltage at the terminals of the transmitting amplifier.

At the receiving station the incoming wave is impressed upon a loop antenna and the receiving set. The resulting detected output is then amplified as indicated and returns to the hybrid coil, passing out on the telephone line without producing a voltage on the speech amplifier of the transmitting set if the hybrid coil balance is perfect.

On the ship, this physical separation of transmitting and receiving set is, of course, not practical, and, as indicated before, transmission and reception take place upon one antenna so arranged that the receiving circuit offers a high impedance to currents of the outgoing frequency and low impedance to the incoming signal. Actually,

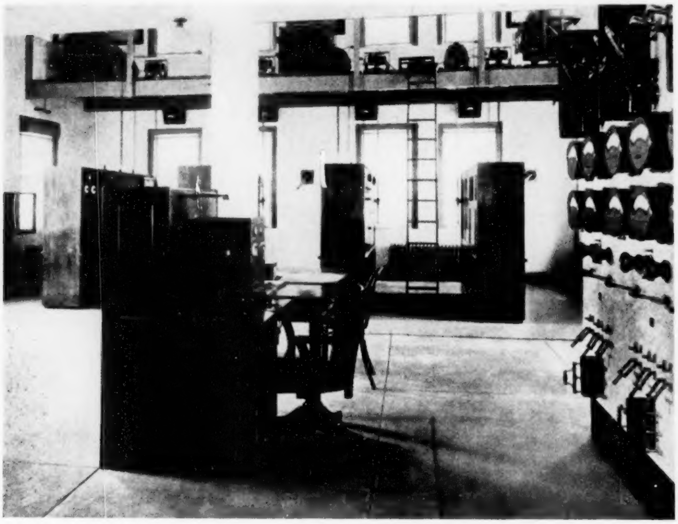


Fig. 3

the outgoing signal is not entirely excluded from the ship's receiver and there is present a side tone of about the magnitude of the incoming signal. This is by no means an undesirable condition and is the one which holds approximately in an ordinary telephone subscriber's instrument. The presence of side tone assures the speaker that his system is functioning properly.

Fig. 3 is a view of the operating room showing the four transmitting units at the back; the power switchboard for supplying the plate circuits of the tube at the right; the telephone switchboard

for the four speech or telegraph channels in the center; and on the gallery above the transmitting units, the coupling coils, loading inductances, and so on, between the sets and the antennas. The motor generator sets capable of supplying as much as five kilowatts at eighteen hundred volts to each of the transmitting sets are located in an adjoining room and controlled from the operating room.

Fig. 4 shows the interior of one of the transmitting units. In the shielded box at the upper right hand corner is the master oscillator which sets the frequency to be used for that particular channel.

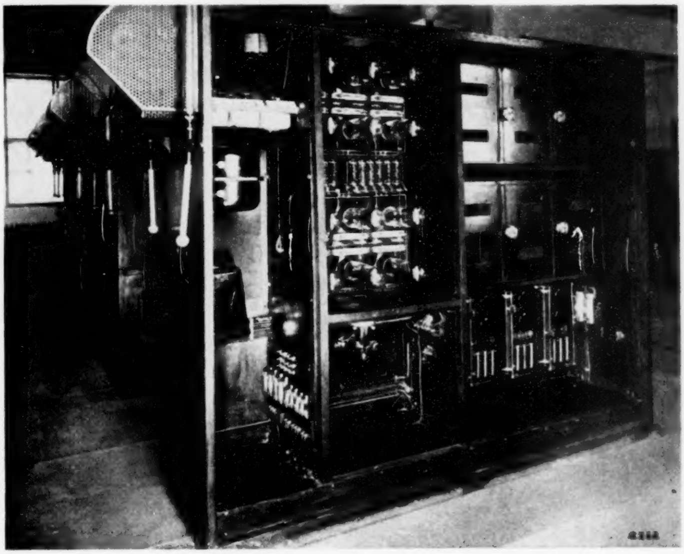


Fig. 4

Next to the left are two more shielded compartments, each of which contains an amplifier. The last stage of this amplifier employs a fifty watt tube. In the larger unshielded compartment are located above, the six radio frequency power amplifier tubes. The reason for introducing two amplifiers between the master oscillator and the power tubes is to prevent any reaction from the attenna circuit back to the master oscillator. By taking this precaution the frequency of the master oscillator never varies more than fifty cycles in eight hundred thousand. The lower set of shielded compartments, at the right, contains the audio frequency telephone amplifiers which supply currents to the six modulator power tubes shown in the lower par

of the open compartment. These two sets of six tubes each are connected together to secure constant current modulation. The output of these twelve tubes is led to terminals on the output of the transmitter unit at the left. To secure cooling in hot weather, a fan is installed below the power tube compartment. In the extreme left compartment are shown choke coils in the power circuits, and at the extreme left on the outside are circuit breakers and two handles for operating the tuning and coupling apparatus in the gallery above.

Fig. 5 shows one set of radio frequency apparatus in this gallery. The two inductometers at the left are for coupling and tuning, and

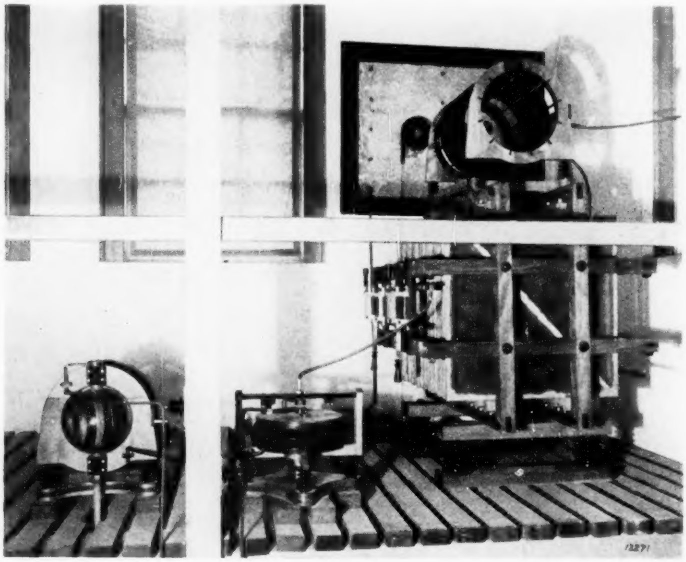


Fig. 5

the large condenser at the right, the plates of which consist of brass frames covered with copper window screen, is inserted in series with the antenna. This capacity together with the inductance immediately above it stiffens the antenna circuit and increases the frequency selectivity to prevent radio frequency interaction between the several antennas.

The telephone switchboard shown in Fig. 6 is a special type of P. B. X. (private branch exchange), constructed to provide the necessary shielding and to include telegraph oscillator, phantom coils, and other special apparatus. This switchboard provides for

four channel telephone or telegraph operation and for the control and monitoring of all channels. In the operation of the system one operator, located at this switchboard, has complete control of the entire transmitting plant. The operating board was especially built

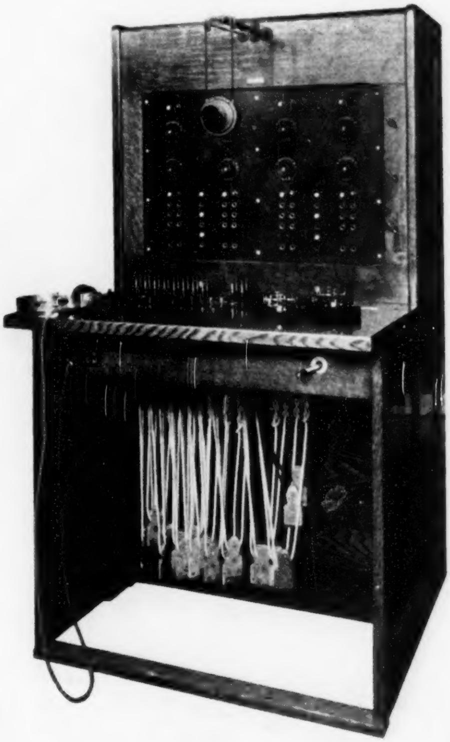


Fig. 6

for the experiments, and altho not the final form contains features which are of interest in that they illustrate well the technique involved in combined wire and radio operation.

The four vertical rows of jacks correspond to the four two-way radio channels. At the top of each row will be seen the dials for controlling amplification. On the apron are telegraph keys, telephone keys, and operating cords. The cord circuits, by being plugged into the jacks, interconnect any one of the New York toll circuits with

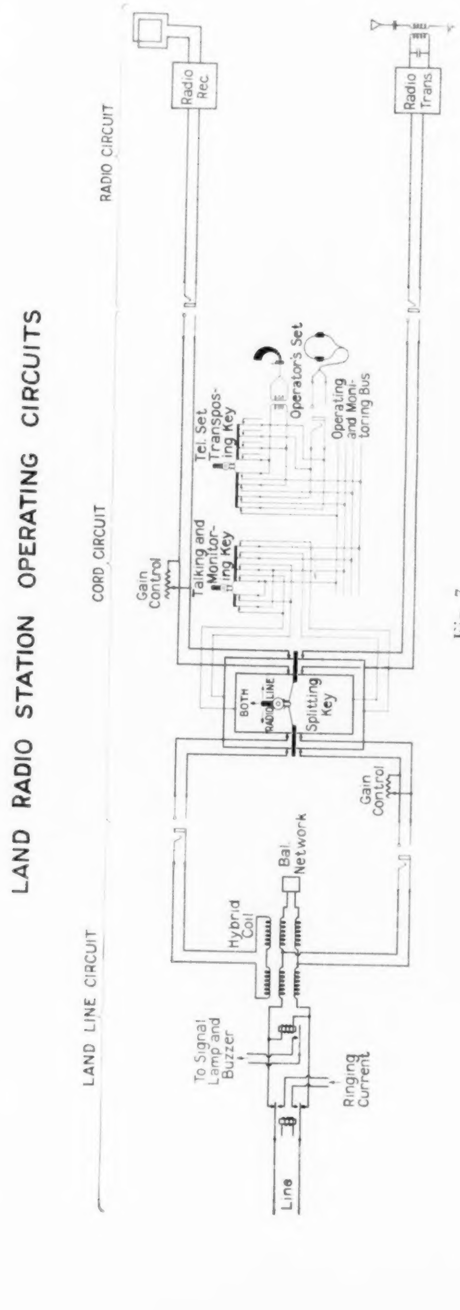

LAND RADIO STATION OPERATING CIRCUITS
LAND LINE CIRCUIT
CORD CIRCUIT
RADIO CIRCUIT
Radio Rec.
Radio Trans.
Gain Control
Talking and Monitoring Key
Tel. Set Transposing Key
Operator's Set
Operating and Monitoring Bus
BOTH
RADIO
LINE
Splitting Key
Gain Control
Bal. Network
Hybrid Coil
To Signal Lamp and Buzzer
Ringing Current
Line

Fig. 7

any one of the four radio circuits. The cord circuits contain the switching keys seen in front, by means of which the radio station operator is enabled to split the circuit and talk either way, connect the circuit thru and bridge on it and talk or monitor. This cord

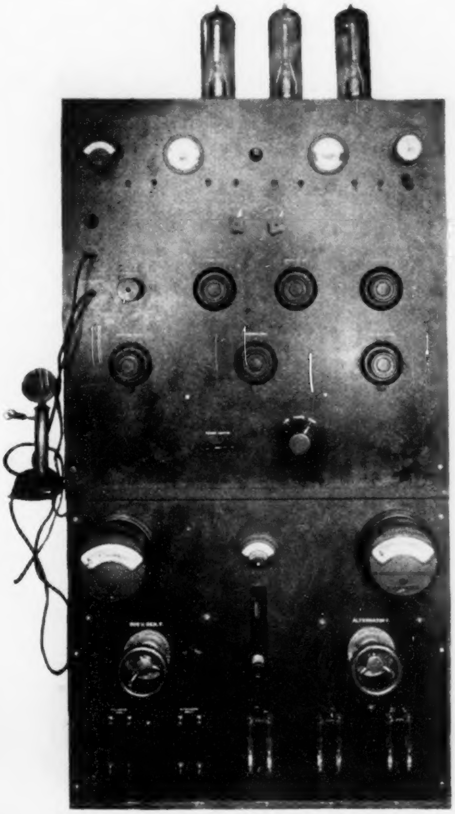


Fig. 8

circuit is shown in Fig. 7 in relation to the rest of the wire-radio junction circuit. It will be seen to be of a four-wire instead of the more usual two-wire type and to comprise in reality two circuits, one for east-bound and the other for west-bound transmission. This

arrangement was used in order to obtain flexibility in the experiments. It enables the circuits to be continued inland as four-wire circuits and permits of the switching operations being carried out with a minimum effect upon the 2-way balance of the transmission system.

The receiving station at Elberon is located on a rented plot of ground and was not built in permanent form, since we did not regard this location as entirely suitable for receiving from the Atlantic. Reception is carried on the four channels by means of four loop antennas operating into four receiving sets. A telephone switchboard similar to that at Deal Beach provides for the connection to the wire system. Of course, two telephone switchboards are not necessary but one was installed at each station in order that we might determine by operating tests whether the control of the system should be from the transmitting or the receiving station.

The receiving sets as finally developed were extremely selective and pass only a band of speech width with a large attenuation outside this band. They will be described in another paper.

Fig. 8 shows a front view of one of the experimental transmitters used aboard ship. The lower half consists of power control apparatus. Three 250 watt tubes are used of which one is a master oscillator, one a power amplifier and one a modulator. The large capacity tube was used as a master oscillator and only a very small part of its output applied to the second power tube. This was done in order to prevent reaction of the antenna system upon the oscillator.

Apparatus of this type was installed on the *Ontario* and *Gloucester* of the Merchants and Miners Line, and operated in conjunction with Deal Beach and Green Harbor. Later another electrically similar set was built by the General Electric Company and was installed and operated by the Radio Corporation on the steamship *America*. This installation is illustrated in Fig. 9.

Part II

OPERATION OF THE COMBINED RADIO AND WIRE SYSTEM

The development work as described in Part I had resulted in establishing an experimental ship-to-shore radio telephone plant of some proportions. This will be seen by reference to the accompanying map of Fig. 10 which gives a picture of the field setting, as it were, of the experimental operations. The experimental plant included:

Two operating shore stations—Deal Beach, New Jersey, and Green Harbor, Massachusetts.

A field experimental station at Cliffwood, New Jersey.
Two ship installations, on the S. S. *Gloucester* and the S. S. *Ontario*. The vessels operated between Boston and Philadelphia or Baltimore

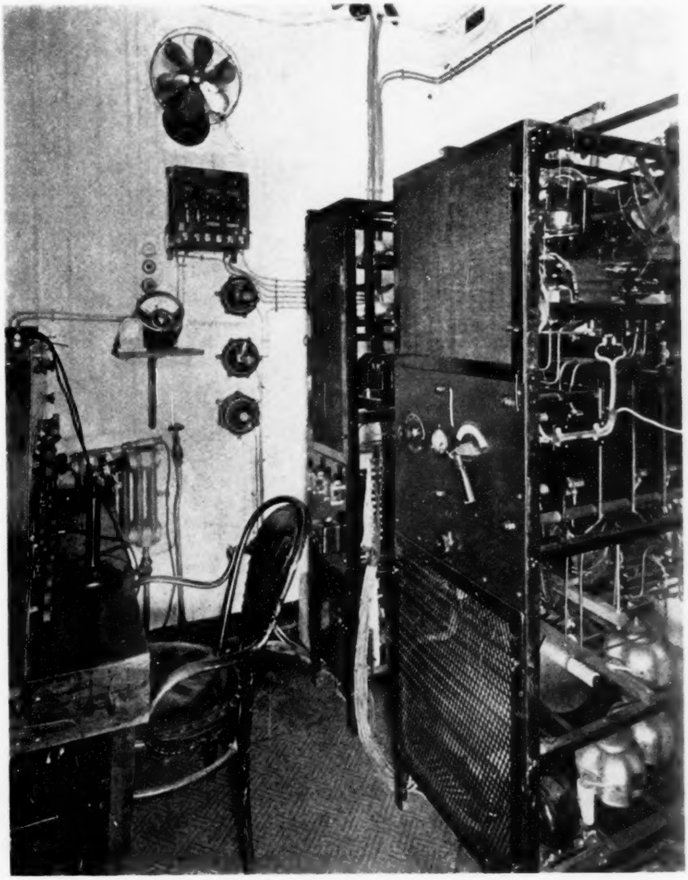


Fig. 9

and took any of several courses, two representative ones of which are as plotted in the figure.

Let us now consider this plant from the communication standpoint and look into its characteristics, first as an electric transmission system, and then as a message handling facility.

Each of the two land stations was tied into its nearest center by wire circuits, Deal Beach to New York and Green Harbor to Boston. We will take for our example the New York-Deal Beach-Ship circuit pictured in Fig. 11 and shown diagrammatically in Fig. 12. This is

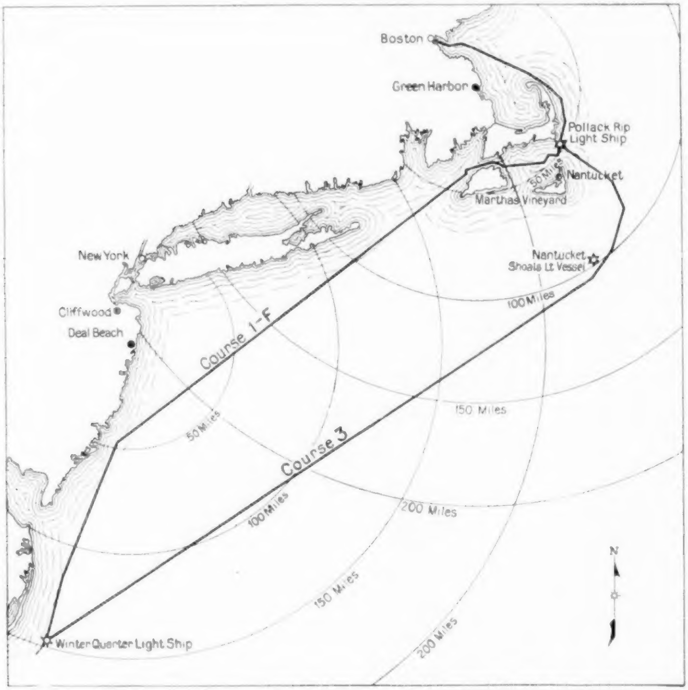


Fig. 10

a combination of wire-radio toll circuit, one end of which terminates on a vessel of variable position and the other end of which is capable of being extended either over a local circuit to a New York subscriber or over a long distance circuit to reach subscribers at more distant inland points.

This communication circuit must fulfil two general requirements. In the first place it must be so constituted electrically as to preserve the feeble voice currents launched upon it by one subscriber so that they be rendered to another person with sufficient volume and fidelity of wave shape as to be readily intelligible. This requires that the circuit be properly engineered as an electrical transmission network.

Secondly, given a circuit capable of talking, it is necessary that this circuit be flexible in use so that it can be put at the disposal of any land line subscriber for connection to a ship at sea, at any time the ship is within range. This requires that the proper switching facilities be provided and brings in operating and traffic problems.

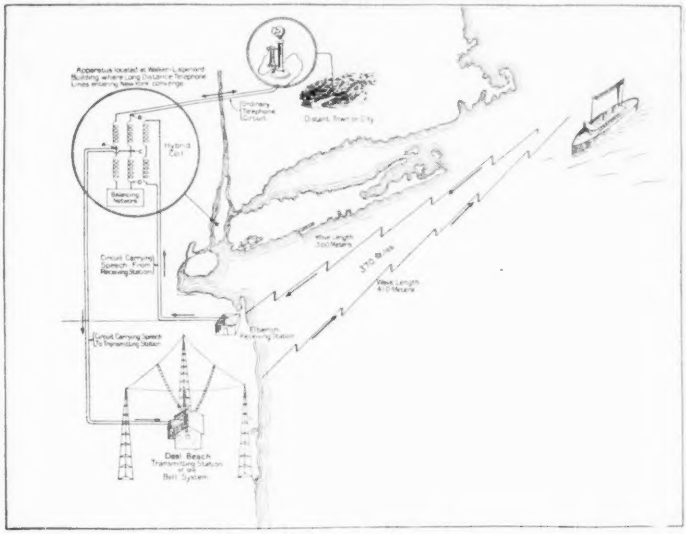


Fig. 11

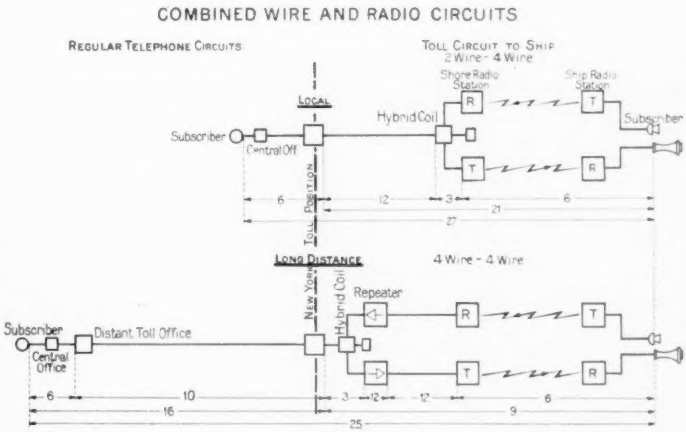


Fig. 12

The Radio-Wire Transmission Circuit

Two types of circuits were used in the experiments in operating between New York and the vessel, as shown schematically in Fig. 12. The radio link is the same for both. Different frequencies are used for transmitting in the two directions so that in the radio link we have the equivalent of two circuits, one for transmitting east and the other for transmitting west. This is the same as the four-wire telephone circuit. In the thru circuit first shown the radio four-wire circuit is brought into a two-wire circuit at the land radio station, making a regular two-wire telephone circuit from the radio station to the New York toll terminal. This circuit was used in the tests for calls local in New York. In the arrangement of the second circuit of Fig. 12, each of the one-way radio channels is extended back to the New York toll office by its own wire line and there joined into a two-wire circuit, making the wire-radio toll line from New York to the ship a four-wire system. This arrangement forms a high grade circuit and is the one used in the experiments for connecting with long distance lines.

Taking the four-wire circuit, the path of the voice currents may be traced thru from one end to the other as follows: The currents which are initiated by the land subscriber, for example, upon arrival at the New York toll office divide in the hybrid coil between the east-bound and the west-bound circuits. The currents are prevented from being propagated over the west-bound branch because of the unilateral nature of the repeater. The currents of the east branch are amplified in the repeater of this circuit in order to make up for the attenuation suffered in the cable circuit, and upon arrival at Deal Beach are amplified to power proportions, modulated upon the radio carrier and radiated into space. Upon being received at the ship end of the radio circuit they are sharply selected in respect to frequency, are again amplified, and delivered to the listener. When the ship subscriber talks, the voice currents are amplified, pass directly into the radio transmitter, are transmitted over the radio link in the usual way, received at the shore station, amplified and sent out over the wire circuit to New York. Here they pass in the reverse direction thru the hybrid coil and divided between the two-wire circuit on the one hand and the balancing network on the other, thus getting back into a regular two-wire telephone circuit.

Interconnection Between Radio and Wire Circuits

It will be well to recall at this point just what it is that makes possible automatic repetition or thru transmission between the wire and radio circuits.

There is both an outgoing and an incoming radio channel. The automatic repetition from the wire to the outgoing radio channel is made possible thru ability to control the transmitter wave power by the voice currents set up at the distant end of the telephone line. It will be recalled that in the early radio telephone art, before the vacuum tube, modulation was effected by the microphone transmitter which required that the talker be present at the radio station. It is, therefore, the electric-control type of modulator such as the vacuum tube, as distinguished from the air-wave control modulator, which permits of the talker being at the far end of a wire circuit. Conversely in the receiving channel, it is the fact that the detecting action yields telephone currents directly, ready for propagation over a wire circuit, that enables the radio channel to be extended to a distant listener.

Thus it is the thermionic tube modulator and detector which have made possible the radio-wire transfer. It is the thermionic tube as a reliable high-quality amplifier, however, that makes the transfer practical; for it is the amplifier which enables the weak voice currents received at the radio station from a land line subscriber to be boosted to power proportions and thus control the considerable radio frequency power required for transmission; and, again, it is the amplifier which enables the extremely weak currents received from the radio link to be so augmented that upon being placed upon a wire circuit, and perhaps being further amplified en route, they may be heard in the regular telephone at the other end.

The other important feature of the radio-wire inter-connection is the junction of the four-wire and the two-wire circuits by means of the hybrid coil and balancing network as shown in Fig. 11. The windings of such a coil are so designed as to establish a sort of Wheatstone bridge circuit. This bridge circuit accomplishes the joining of the regular two-wire telephone circuit with the sending radio channel on the one hand and the receiving radio channel on the other, while still maintaining an electrical separation between the two radio channels. It is really, therefore, the connecting link between the two-wire type of circuit of the telephone plant and the four-wire circuit of the radio link. The hybrid coil type of circuit is taken from the telephone repeater and carrier current art.[2] The radio receiving circuit corresponds to the generator branch of the Wheatstone bridge, and radio transmitting circuit to the detector branch. The

[2] "Telephonic Repeaters," B. Gherardi and F. B. Jewett, *Journal of American Institute of Electrical Engineers*, pages 1255–1395, November, 1919.

"Carrier Current Telephony and Telegraphy," E. H. Colpitts and O. B. Blackwell, *Journal of American Institute of Electrical Engineers*, pages 205–300, February, 1921.

two-wire telephone line corresponds to the "X" arm of the bridge and the balancing artificial line to the "Y" arm. The ratio arms are in effect formed by the windings of the hybrid coil.

Speech Received from Ship Re-Transmitted from Shore Station

Now this junction circuit always has some unbalance because it is obviously impossible to maintain a perfect symmetry between the telephone line and the balancing network. Especially is this true where the telephone line is a type not designed for repeater operation and is switched at its terminal to any of a number of lines of different impedances, as was the case with the circuit used in the tests.

This unbalance between the line and its balancing network will be seen to permit some of the speech-current received over the radio link to get across into the transmitting circuit, to modulate the shore station carrier and to get out into the ether again on the transmitting wave length. As a matter of fact during the experiments the unbalance was sometimes such as to permit of fairly strong transmission around back thru the shore transmitter, so that incoming speech was repeated out thru the shore station transmitter in amplified form. This enabled listeners in the vicinity of New York to hear the conversation originating on the ship almost as well as that originating on land, and they naturally thought that they were picking up the ship's radio transmission directly, whereas they were actually overhearing the re-transmission of the shore-station's reception.

The Thru Circuit as a Repeated Telephone Circuit

This re-transmission makes all the more evident the true role of the shore station, namely, that of a large telephone repeater between two sections of line, the one a land line and the other a "space" line, and functioning also to convert between the voice frequencies of one section and the radio frequencies of the other. As such, we can consider the over-all circuit from a transmission standpoint much as we do long distance repeated telephone circuits.

Now one of the most important transmission considerations in such a long distance circuit is that of how the amplification is applied in relation to the losses in the circuit. This question of amplification is particularly important in the case of combination radio-wire systems, because the radio circuit possesses inherently large transmission losses and requires correspondingly large amplification. The

necessary large amplification is supplied at both ends of the radio link, partly in the transmitting station where the voice currents are amplified up to power proportions and partly at the receiving end where the amplification is likewise large altho at small power. In the radio telephone circuits which were operated in the experimental work the power in the sending antenna to that in the receiving antenna is in the ratio of roughly 10^{10}. This requires amplification which was distributed somewhere near equally between the sending and the receiving ends. It has been found convenient to express such transmission losses in terms of a power ratio using $10^{0.1}$ as a unit.[3] Thus the above antenna to antenna power ratio would correspond to 100 of such units.

Circuit Transmission Equivalents

It is necessary that the amplification of such a circuit be sufficient to offset very closely the loss, in order that the net loss be small. Actually, in the tests, the radio portion of the circuit was worked with a net transmission loss of about six units, meaning that at least 95 per cent. of the radio over-all circuit losses were wiped out. This means that if a change of say 10 per cent. occurs in the amplification, or in the ether loss as by fading or movement of the vessel, the circuit equivalent will be greatly affected—changed by about 200 per cent. The difficulty of maintaining the ship circuit stable will therefore be appreciated.

Fig. 12 shows the transmission loss (of six units) obtained for the radio link during the tests and also the other losses which are in the wire portion of the combination system. The distribution of losses in the first circuit will be noted to be approximately as follows:

6 units in the radio link.
3 units in the hybrid coil—balancing network.
12 units in the wire circuit to New York.
6 units from the New York central office to the subscriber.

[3] The unit used in this paper is one which has been found convenient in which expressing the transmission loss or gain of a circuit. One unit is taken as that power ratio which is equal to $10^{0.1}$. Thus, if the attenuation or amplification of a circuit is one unit the power at the two ends are in ratio of $10^{0.1}$; if ten units, in the ratio of $10^{10\times0.1}$ or 10; twenty units would therefore have a power ratio of 100, and so on. The advantage of using a power ratio instead of a current ratio is that it is undependent of the impedances of the two portions of the circuit considered. The advantage of expressing the power ratio as an exponent is that on account of the exponential nature of attenuation it enables the net transmission efficiency of a system to be readily derived by algebraically summing up the individual losses and gains. This unit has been selected as more suitable for general use in expressing transmission efficiencies than the 800 cycle "mile of standard cable" which has sometimes been made. The ratio between these two units is as follows: 1 mile of standard cable equals 0.95 units as used in this paper. $\frac{P_1}{P_2} = 10^{0.1} = \frac{1}{0.95}\epsilon^{(0.109)}$ units.

This makes a total loss between subscribers of 27 units which is satisfactory for a good talk under fairly quiet conditions. This equivalent was usually realizable under the conditions of the test and the majority of the calls put thru from local stations in New York with the ship 100–200 miles (160–320 km.) out were successful despite occasional spark interference in the radio circuit.

The transmission loss is, however, too high in such a circuit to enable it to be extended inland over long distance wire circuits. If this is attempted, two limitations come into play. In the first place, the volume of the talk becomes too weak. If the call were extended over a toll circuit having a 10-unit equivalent, for example, the overall equivalent would become something like 37 units, which is excessive. This could be overcome to some extent by a cord circuit repeater at New York. A second limitation which existed in the experimental set-up resided in the unbalance between the line and the balancing network at the radio station. This unbalance permitted currents received over the radio link to be fed back thru the radio transmitter of the land station, as described above. These fed-back currents overload the radio transmitter if they are large compared with the currents being supplied to the radio transmitter from the shore subscriber. In other words, if there is sufficient amplification in the shore transmitter to enable very weak voice currents arriving over a line of high equivalent to load the transmitter fully, then the transmitter is likely to be overloaded by currents which get through the hybrid coil from the associated radio receiver. For these reasons, the two-wire-four-wire circuit of Fig. 12 is not good enough for extension over long distance circuits.

The four-wire type of circuit which is suitable for long distance land line connections is shown in the second diagram with representative transmission equivalents. A brief comparison of the two-wire and the four-wire circuits will make it evident why the four-wire circuit gives the better equivalent. It enables the land line loss between the radio station and the toll center to be more or less wiped out, thus in effect placing the radio station electrically at the toll center. Another way to express the situation is this: regard the hybrid coil unbalance as the limiting factor, then assume that, while holding to a given unbalance, the four-wire circuit (the loss in which can be largely wiped out by one-way amplifiers) is extended inland. The length of the remaining two-wire line back to the land subscriber is thereby decreased and the ratio of the current received at the radio station over the line as compared with that transmitted across the hybrid coil thru unbalance is increased. It will be observed

that with the circuit conditions as illustrated, the over-all equivalent between, say, a Chicago subscriber and a ship, including a 10-unit toll circuit loss, is approximately 25 units, which should give a good talk.

Power Levels and Interference

It is necessary that the magnitude of stray currents be so kept down in comparison to the transmission currents thruout the system as to obviate noise interference with telephone conversation. This requirement is particularly difficult of realization in the radio link because of static and, especially in the vicinity of New York, interference from spark telegraph stations. It is, of course, this interference, caused by the presence in the ether, on the wave length band being used, of extraneous wave components, which sets the actual range limit of the radio link. Actually it was found that in transmitting on about 400 meters in the vicinity of New York the receiving field strength could not be permitted to go on the average below about 200 micro-volts per meter, and even then the spark situation is so bad in the present art as to give periods of prohibitive interference. In less congested zones along the coast to the north, probably lower field strengths could be permitted.

Transmission Variations in Radio Circuit

One of the outstanding transmission characteristics of a ship-to-shore radio telephone system is the variation which the attenuation of the radio link undergoes as a result of the movement of the vessel. In order to determine the magnitude of these variations, a series of measurements were made of the telephone transmission over the radio circuit as the vessel proceeded on her course.

The method of making these measurements is shown schematically in Fig. 13. Take for example the case of measuring a one-way circuit as distinguished from a circuit looped back. A 1,000-cycle current of predetermined power of the order of one miliwatt is impressed upon the input circuit of the radio transmitter. This tone is received in the output of the distant radio receiver. There it is passed to the measuring apparatus where it is amplified, rectified, and made to operate an indicating instrument. The receiving end measuring apparatus is then switched to a local source of 1,000-cycle current giving the same power as was applied to the transmitter at the sending end. The proportion of this power which enters the measuring apparatus is then varied by a variable network calibrated in power ratio

loss (it was actually in miles of standard cable) until the indicator reading is the same as that obtained from the radio receiver. The setting of the variable network then indicates the transmission loss of the circuit. The method is similar to that developed for measuring the transmission loss of telephone circuits.

These measurements were used for the purpose of enabling the radio link to be worked to a contsant transmission equivalent—in this case, of about six units. The procedure for so doing was as follows: When the vessel was first picked up, the receiving amplifications in both east and west channels were adjusted until the measurements showed a transmission equivalent of 6 units. Then as the vessel proceeded on her course the transmission equivalent was

MEASUREMENTS OF TELEPHONE TRANSMISSION OF RADIO CIRCUITS

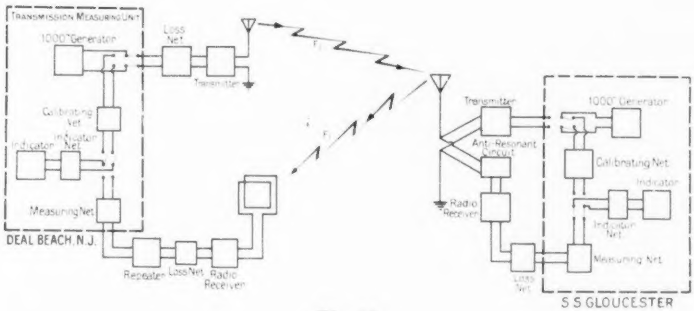


Fig. 13

measured at intervals of an hour or less and the receiving amplification readjusted to hold the desired six-unit equivalent. In this way the talking efficiency of the radio link was kept constant.

The total amplification change which had been made from the beginning of a run up to any one time gives a measure of the change which has occurred in the transmission loss of the radio circuit. By plotting this change in relation to the time of day and in turn to the varying distance between the two stations, an interesting curve results which shows the manner in which the progress of the vessel affects the transmission of the radio circuit. In Fig. 14, the time of day is plotted horizontally, distance is plotted vertically on the right and amplification vertically on the left. The zero amplification reference is the amplification in the circuit which gives a six-unit equivalent at the time the vessel is first picked up.

Curve A shows the manner in which the distance between shore-station and ship varied with time of day as on her south-bound course the ship approached the shore station from the northeast and drew away again to the south. The vessel in this case was on her inshore course along the coast as shown in Fig. 10 above. Curve B shows the manner in which the receiving amplification on shore had to be changed with time of day to keep the ship-to-shore transmission constant, first decreased as the ship came closer and then increased as she drew away again to the south. Curve C is for reception on ship and shows the manner in which the ships receiving amplification

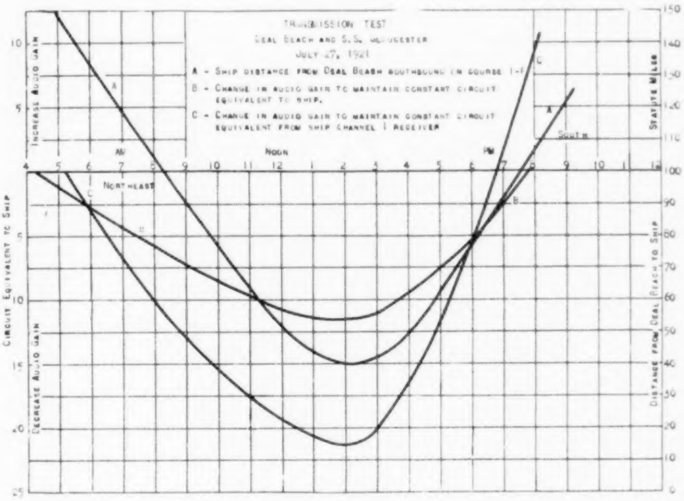


Fig. 14

had to be varied to keep the shore-to-ship transmission constant. Altho these are not measurements of any absolute quantity such as field-strength, they are practical measurements of actual talking circuits and as such include the effect thereon of apparatus adjustments as well as the "ether" conditions and therefore are of value in determining the conditions to be met in maintaining the over-all system in operation. The difference between curves B and C, for example, are in part the result of the difference in adjustment given to the detector tubes of the two circuits. The load on the detectors was kept practically constant by adjustable radio-frequency amplification, so that any overloading was approximately constant thruout

the measurements and had the effect of minimizing the variation of circuits equivalent with distance.

The curve of Figs. 15 and 16 are obtained by taking the data of Fig. 14 and plotting the variation of amplification with separation between stations. The minimum point is arbitrarily chosen. These

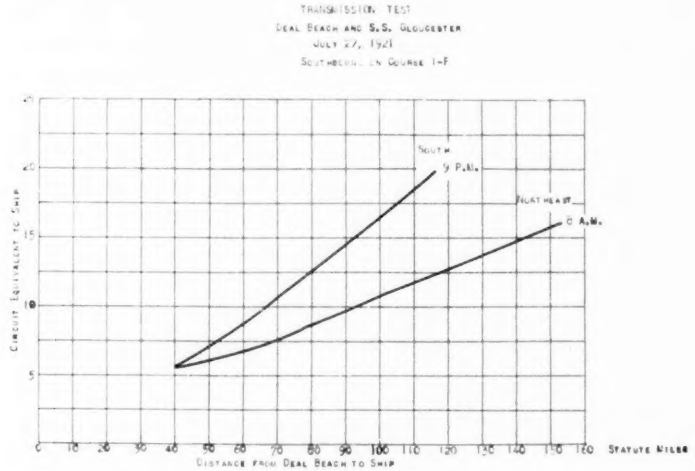


Fig. 15

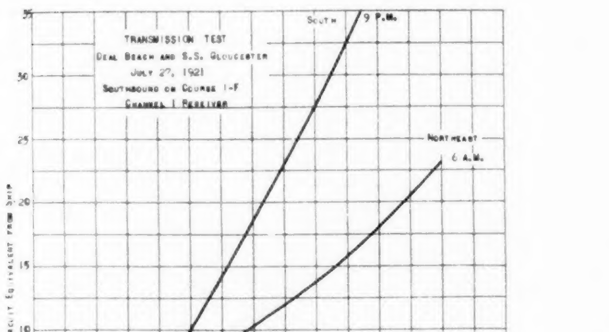


Fig. 16

curves show up rather strikingly the fact that the transmission to the south is much poorer than that to the northeast of Deal Beach. This result agrees with experience since greater difficulty is usually encountered in communicating with a vessel when the line of transmission is along the coast than when it is straight out to sea. The larger transmission loss may be due to shore absorption or, possibly, to refraction of waves by the electrical discontinuity represented by the coast line. This high attenuation effect was observed rather uniformly in all of the measurements, and would itself form an interesting subject of investigation, using more absolute methods of measurement.

Fig. 17 shows the circuit-equivalent-distance characteristic for the off-shore course taken by the vessel. The curve is generally similar to that for the near-in course.

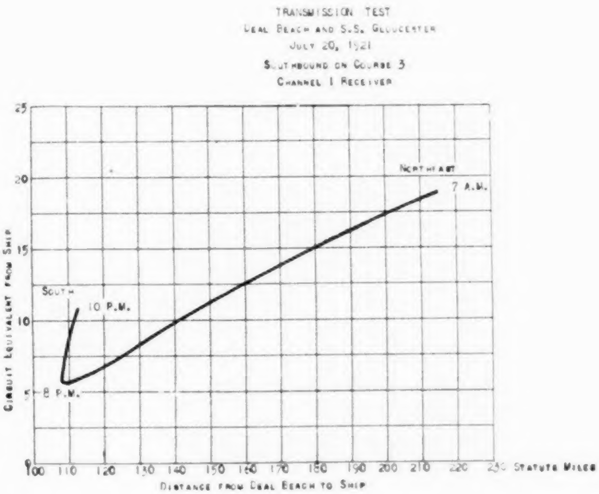


Fig. 17

The rate at which the transmission of the circuit varies with change in distance is an important matter in operation, since it determines the frequency with which the amplification in the circuit must be readjusted to keep the equivalent constant. In accordance with the Austin-Cohen formula, the equivalent should vary at the rate of about 0.25 units per statute mile (1.6 km.) with the vessel 100 miles (160 km.) out, assuming a square law detector. In the worst case found, the circuit equivalent varied at the rate of 0.5 unit per statute

mile. This was one of the cases where the vessel was well south and close in shore and, of course, represents very poor transmission. It means that with the vessel traveling as slowly as ten miles (16 km.) per hour, a transmission change of about five units per hour will occur. In telephone practice it is desirable to keep the transmission equivalent constant to within two or three units, so that this condition would require re-adjusting the amplification as often as every half hour.

These curves show the necessity for so designing the receiving set as to be able readily to obtain a wide variation in amplification in order to accommodate the changes in transmission efficiency. Under the conditions of the test, it was found that a variation of 40 units amplification is necessary in order to carry the vessel from a range of 40 up to 200 statute miles (64 to 320 km.). In order to gain such control, it is desirable to switch stages of amplification into and out of the circuits and to obtain closer adjustment by means of networks of variable loss.

When we speak of holding the circuit to a constant transmission equivalent thruout a wide variation in the position of the ship, we do not mean that the circuit for 200 miles (320 km.), for example, is as good as that for 40 miles (64 km.). The field strength received over the longer circuit is, of course, very much weaker than that received over the shorter one, and is subject to correspondingly more interference. For a transmitting power of about one kilowatt used in the tests, it was found that the circuit was rather consistently good up to 100 miles (160 km.) or so. During summer daylight condition, it was only fairly good at around 150 miles (240 km.), and at 200 miles (320 km.) was subject to so much interference that its insurance against interruption in service was small. Under more favorable conditions, particularly as at night, in the winter time, connections could be established over very much greater distances, but not reliably.

Field Strength Measurements

The circuit transmission measurements described above were supplemented during the latter end of the experiments by measurements of a more fundamental nature, namely, of the received field strength. These measurements represent the application to the ship-to-shore development program of what was really another investigation—that of the development of methods and apparatus for measuring field strengths at these relatively short wave lengths as well as for longer wave lengths. The method and means employed will not be de-

scribed, since they will be the subject of another paper to be given before The Institute of Radio Engineers shortly by the engineers immediately responsible for this development. Field strength measurements will be discussed, therefore, only in so far as the results obtained apply to the ship-to-shore type of system.

The first measurements which were made are given in Fig. 18. They show the variation in the strength of field received from the S. S. *Gloucester* during the same trip as the circuit measurements of Fig. 14 were taken. The fact that the transmission is considerably better to the northeast from Deal Beach than when the vessel is south is rather strikingly in evidence. The dash-line curve is for the Austin-

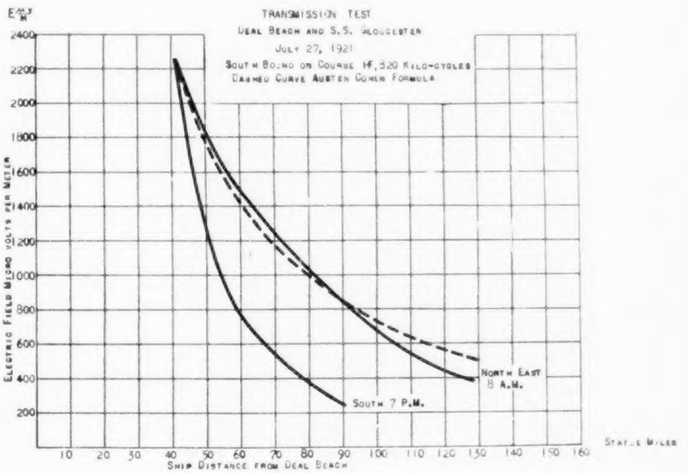


Fig. 18

Cohen formula and shows that the field strength versus distance relation checks that formula for these relatively short wave lengths when the path of transmission is practically entirely over sea. The absorption coefficient is obviously greater when the path of transmission has a relatively large component skirting along the coast line. The curve shows also that the vessel was picked up and the circuit "made" on this day at a field strength of about 400 micro-volts per meter, then increased as the vessel came nearer and passed the land station to a value of 2,200, and that the circuit was "broken" on about 200 micro-volts. The results of another measurement made at the off-shore course of the vessel are given in Fig. 19. A larger

portion of this curve is for straight-out-to-sea transmission where the attenuation law is seen to be normal.

In the field strength measurements made during the ship-to-shore development, those of the S. S. *America* en route across the Atlantic are especially illuminating. The results are given in Fig. 20. The vessel was in-bound so that the curve develops from the right to the left, altho the effect is just the same as if it developed in the reverse direction with the vessel out-bound as was shown by another set of measurements which gave generally similar results. The actual measurement results are indicated by the points and by the con-

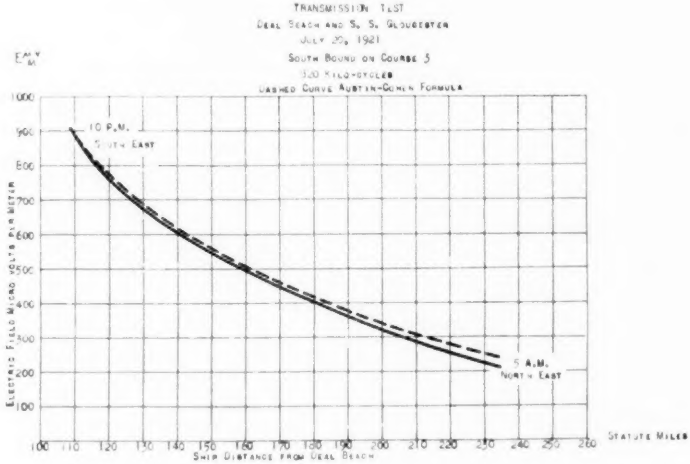


Fig. 19

necting heavy line curve. The light curve A is a plot of the Austin-Cohen formula, the absorption term of which is for daylight transmission over water. The light curve B is a plot of the simple inverse-with-distance law without any absorption term.

This curve shows up the following important factors:

(1) The enormous variation between day and night in the received field strength which occurs at distances of the order of 1,000 miles (1,600 km.) using wave lengths of 350 to 400 meters as now employed in broadcast transmission. The curve shows night to day variations of the order of 100:1 or a power ratio of 10,000:1. This means, for example, that it would require 10,000 times more power to "get thru" as well during the day

as during the best times at night. These enormous day to night fluctuations are now familiar to broadcast listeners. This curve shows the impossibility of giving continuous ship-to-shore telephone service at these relatively short wave lengths for distances as great as 1,000 miles (1,600 km.). For such distances much longer wave lengths will be required, as well as more sending power.

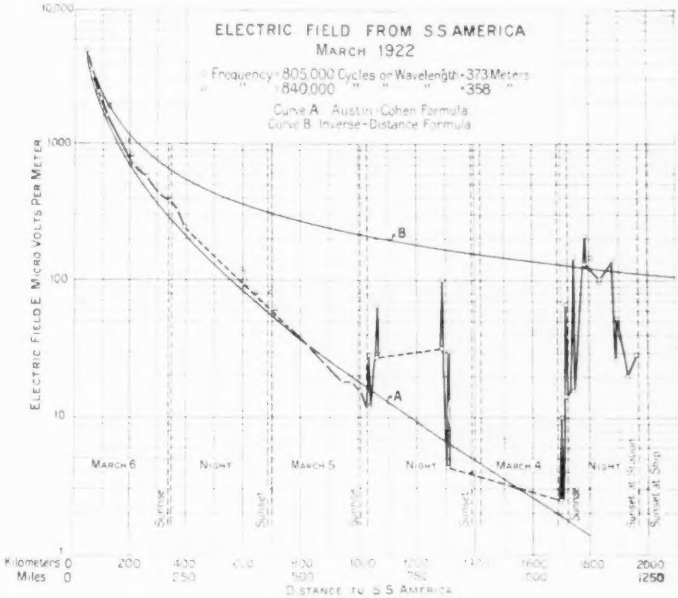


Fig. 20

(2) The wide fluctuations which occur thruout the night period. Altho smaller than the day to night fluctuations, their effect upon transmission is still very large. The fluctuations during the third night out, for example, are as much as 10:1 in field strength or 100:1 in power, or about 20 of the power ratio units we have used above. In view of the rapidity with which these fluctuations occur — within a very few minutes — it is practically impossible to maintain a circuit under these conditions satisfactory for regular telephone service.

(3) The most interesting thing to observe is that the fluctuations tend to fall within the two curves A and B. The day trans-

mission is a pretty definite proposition, following closely the Austin-Cohen formula. The night transmission appears in the nature of a "bob-up" from the day condition but seems to be limited in the extent of its "come-back" by the loss imposed by the simple inverse-with-distance law. The fact that the difference between curves A and B is entirely one of absorption suggests that the very large and rapid night fluctuations, which are now so well known to broadcast listeners, may be explained in large part if not in whole, by variations in atmospheric absorption.

Occasional Long Distance Transmissions

Many of the long distance records which have been made on short waves and low power can be accounted for simply on this basis—that the absorption which ordinarily obtains during daylight has been temporarily wiped out. The way in which it is possible for the range to "open up" tremendously under exceptionally favorable conditions will be seen from this: Referring to Fig. 20, assume that the normal daylight range between S. S. *America* and New York was 250 miles (400 km.) as fixed by a limit taken as 200 micro-volts per meter. Then, at night, this same field strength may be delivered over a distance of about 700 miles (1,100 km.) if the absorption is wiped out in accordance with curve B.

Furthermore, so favorable is the simple spreading-out law at such distances, that the field strength is only halved in going another 700 miles to 1,400 miles (2,200 km.) and only halved again in doubling this distance to 2,800 miles (4,500 km.), and so on. In other words under no-absorption conditions, by increasing the receiving radio frequency amplification by a current ratio of only $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$, or about 12 power radio units, the range of transmission may be increased from the reliable daylight range of 250 miles (400 km.) to a possible night range of ten times this distance. It is therefore seen that many if not all of the long distance transmissions which have been realized for short periods of time probably can be explained simply on the basis of there having occurred an exceptional clearing up of absorption at a time of unusually favorable interference conditions.

Setting Up and Operating Combined Radio-Wire Circuits

The operating problems presented by the combination wire-radio telephone system are more difficult than those involved in the operation of either a straight telephone toll line on the one hand or the

ordinary radio-telegraph circuit on the other. In regular long distance telephone circuits we have a fixed type of system which is maintained continually in good talking condition and the operators turn the terminals over to the use of the subscribers themselves. On the other hand in a radio telegraph circuit operating between land and vessel the circuit is kept entirely in the hands of skilled operators who have access to the apparatus and who handle the traffic directly between themselves. In no case before have we had the requirement of taking a radio link of varying length, building it up as occasion

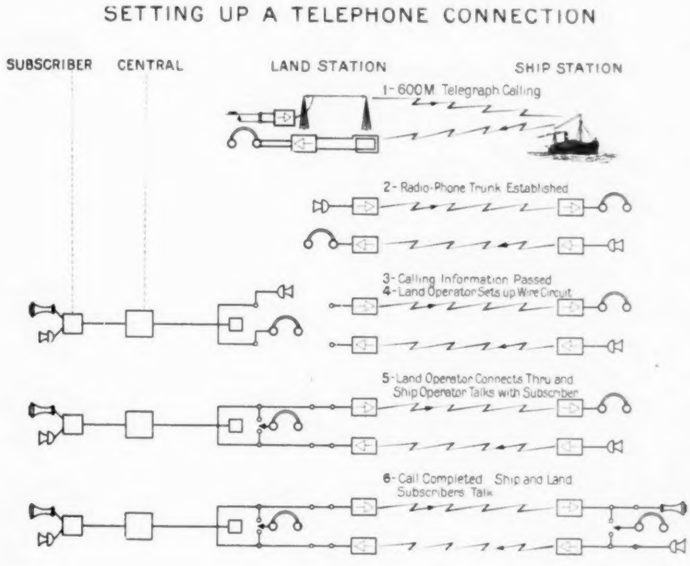


Fig. 21

requires with wire circuits and, upon call, putting the combined system at the disposal of people experienced only in the use of the regular telephone. The technical difficulties of the combination system, together with the necessity of coming as close as possible to meeting telephone standards in the quality of talk given, greatly reduce the length of the radio link which can be used for a service as compared with those distances which can be spanned for short periods of time under the most favorable conditions. The effect which the requirement of reliability has in reducing the range of transmission will be appreciated from the discussion of field strength measurements.

There are various ways in which the combination circuit can be set up and operated and it will take further experience before the most satisfactory arrangement is determined upon. In order to explain the operation generally, however, we will describe how the circuit was actually set up during the tests. Take the case of a call originating on the vessel; then the procedure is as illustrated in Fig. 21, namely:

1. When the ship comes within range she calls the land station by telegraph on 600 meters, and informs the land station of her message business.
2. The land station then assigns a pair of telephone channels to which both stations switch over and the circuit tested out for talking. In case of important long distance land line connection, this test may involve circuit transmission measurements.
3. The ship operator then passes to the land operator by voice (or by tone modulation telegraph) the information as to the connection desired.
4. The land operator then tells the ship operator to stand by while he switches to the wire circuit and passes the call to the telephone central, who in the case of a local call is a local operator (actually, for this case, she was the operator on the Cortlandt Official Board of the American Telephone and Telegraph Company at New York; or in the case of a long distance call is a toll operator.
5. The land line connection is made in the usual way and the shore station radio operator greets the land line subscribers.
6. The shore station operator then joins together the land line and the radio link thus connecting the land subscriber with the ship operator, who proceeds to tell the subscriber that this is the steamship so and so and that Mr. Blank wishes to talk with him. While this is going on, the land operator is monitoring on the circuit and makes such final adjustments of the amplification as may be necessary.
7. The ship operator then summons the ship subscriber and the latter takes up the conversation.

The handling of calls originated by the land line subscriber presents a more difficult operating problem because of the uncertainty as to the radio link—it not being known whether it can be established and, if so, as to how long a wait will be involved in getting the connection. During the tests, most of the calls originated in New York

area. For these cases the land subscriber was connected to the Deal Beach station and there the call was put thru directly to the ship in case the radio telephone circuit was available. When not available, information as to the call was recorded by the radio operator and the telephone circuit released for the time being. The call was then completed by first setting up the radio link and then calling back the initiating subscriber. It is obvious that the giving of commercial service will involve: first, the ascertaining of whether or not the vessel is within range; second, the "lining up" of the radio link preparatory to the thru connection; and third, the building up of the land line connection back to the calling subscriber and the making of the thru connection. Many detailed variations are possible in the procedure and the determination of the best operating methods will have to await upon experience obtained in actually giving service.

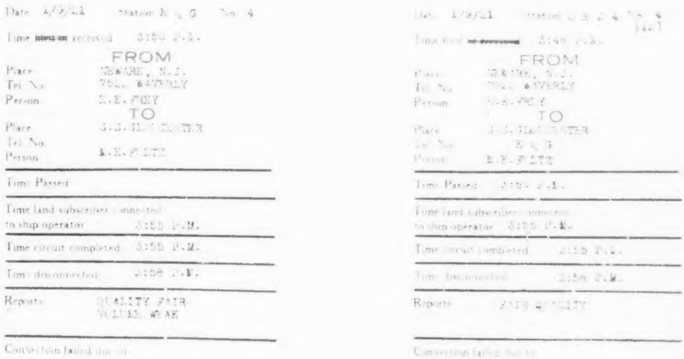


Fig. 22

In order to become familiar with the problems involved in maintaining the ship-to-shore system in operation, a series of operating tests were carried out for a period of about three months, starting in January, 1921, and operating between Deal Beach station and the S. S. *Gloucester*. In accordance with a pre-arranged schedule (unknown to the engineer-operators), calls were entered by a considerable number of Bell System engineers in the vicinity of New York, and calls were initiated from the vessels also by the opening of sealed envelopes carrying instructions to call one or more parties on shore. Fig. 22 is a facsimile of the message form or "ticket" used in the operating tests. The table below gives a representative record sheet recording the calls which were made on a partciular day and also the time which elapsed in putting each one thru. These data, of course,

are not especially representative of what can be done with a system after it has begun smoothly in commercial service, but are interesting in giving a general idea of the way the system worked and in showing that calls were successfully put thru in a reasonably short time. In the aggregate a large number of calls were made, and as a result the system was put to a fairly severe operating test. It was found, as was to be expected, that the time required to put thru the

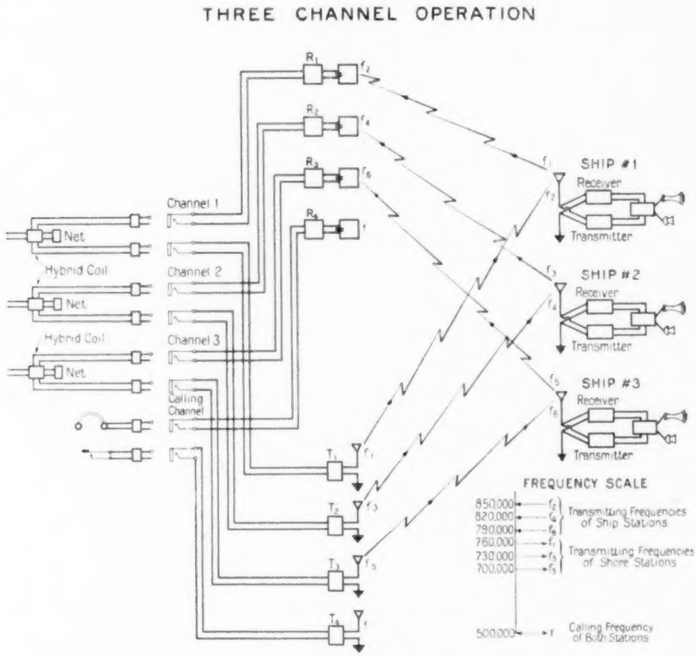


Fig. 23

radio connections to the vessels is large in comparison with the connecting time on the wire lines, and will require that precaution to be taken in the operating routine to minimize the time during which wire circuits are held up pending connection with the radio link. (The wire circuits used in the tests appeared in New York on a busy P.B.X. (private branch exchange) and did not receive the operating attention that they would in regular service.) But even tho the radio holding time was larger than the usual wire time, it is in itself rather surprisingly good considering the difficulties which attended this maiden

operation of telephone circuits to ships and must be regarded as full of promise for the extension of telephone service to the highways of the sea.

Tests of Three-Channel Operation

Of course, any comprehensive ship-to-shore radio telephone system must be capable of establishing a number of telephone connections from a common land station to a number of ships. The Deal Beach experiments, therefore, had as one of their objectives the trying out of multi-channel operation. These tests were conducted during the fall of 1920 and thru January, 1921, with the S. S. *Gloucester* and the S. S. *Ontario*. A third boat was simulated by a small-power experimental set installed at the Cliffwood, New Jersey, experimental station. The three channel operation is illustrated diagrammatically in Fig. 23, which also shows the scheme of frequencies. The three channels transmitting from Deal Beach were grouped in one frequency range, spaced 30,000 cycles apart. The frequencies transmitted from the ships and received at Deal Beach were grouped in another frequency range removed 30,000 cycles from the first and having frequency intervals likewise of 30,000 cycles. Transmitting and receiving channels differing by 90,000 cycles were paired in the manner indicated to form two-way circuits. While it is possible to squeeze channels together more closely than this, it was not desired on the experiments to go to the limit of frequency squeezing, particularly because of the severe selectivity requirements imposed upon vessel equipment. These frequencies represented a fair balance between technical perfection on the one hand and practically realizable conditions on the other. It will be seen that the set-up was really a four-channel system, with the fourth channel used on 600 meters for calling purposes. Under these conditions three conversations were carried on successfully from the single land station, two to actual ships and one to a "dummy" ship at the Cliffwood experimental station.

Equipping of S. S. "America"

The primary development work of the ship-to-shore system was carried out, as described above, in conjunction with coastal vessels. Such vessels were chosen because the rapidity of their turn-round gave much more frequent test periods than could be obtained by means of vessels pursuing a longer route. It remained, however, to equip a trans-oceanic vessel and connect her into the telephone system.

In 1921, the development tests of ship-to-shore telephony were extended to include the General Electric Company and the Radio

Corporation of America. The engineers of these companies built a ship set similar to that developed in the work described above, but of a more commercial design, and installed in on the S. S. *America* in January, 1922.[4] During the succeeding few months, tests were made between the S. S. *America* and the shore, and on a number of these trips connections were put up to various interested parties around New York when the ship was within about 300 miles (480 km.) of the Deal Beach station. Of course, the *America* was carried out much farther than this at night, but the circuits were not sufficiently reliable to be used in connection with the land lines as will be appreciated from the field strength measurements given above.

A photograph of a portion of the installation on the S. S. *America* is reproduced in Fig. 9 above. The talking tests made with the *America* were the occasion of much interest on the part of the listeners-in, and several of the demonstrations which were given the subject of newspaper accounts and need not be described. The more technical phases of the tests with the S. S. *America* are (a) the field strength measurements, and (b) the simultaneous telegraph tests discussed below.

Simultaneous Telephone and Telegraph Operation Between Ship and Shore

During the experiments with the Steamships *Gloucester* and *Ontario*, the radio telephone transmissions were carried on alternately with the conduct of the regular radio telegraph service of the vessels. Simultaneous operation was impossible because the vessels were equipped with spark transmitters. While this arrangement of having to switch between either telephone and telegraph operation is permissible for small vessels where the communication load is light, it is, of course, not satisfactory for large trans-oceanic vessels where the message business may be such as to require practically continuous operation on the part of both services.

Recognizing, therefore, that one of the problems attending the successful application of radio telephony to large vessels is that of simultaneous telephone and telegraph operation, tests of such transmission were conducted in co-operation with the Radio Corporation of America from the S. S. *America*. These were made during February and March of 1922. On the land ends, the two radio circuits terminated at different stations, the telegraph at the Bush Ter-

[4] See article "Duplex Radio Telephone Transmitter," by Baker and Byrnes, *General Electrical Review*, August, 1922.

minal, New York City station of the Radio Corporation, and the telephone at our Deal Beach, New Jersey, station. The telegraph transmitter was of the continuous wave, vacuum tube type manufactured by the General Electric Company. The telephone and telegraph sets used individual antennas on the ship.

Altho certain apparatus difficulties were experienced aboard the vessel because of the short notice at which the tests were made, nevertheless the tests were successful and demonstrated that a telephone set can be made to operate simultaneously with a suitable C. W. (continuous wave) telegraph transmitter. The final solution of this problem of simultaneous operation, however, will undoubtedly require further work in co-ordinating the two types of systems, in order to permit them to be operated on wave lengths relatively close together. During the tests, the wave lengths were widely different, the telegraph operating on about 2,100 meters and the telephone on about 375 meters. The work done at Deal Beach in the development of multiplex telephone operation, where three telephone channels were operated in the vicinity of 400 meters and a fourth channel was operated for telegraphy at 600 meters, demonstrates that it should be feasible to operate telephone and telegraph channels simultaneously on closely adjacent wave length bands. However, in determining wave length allocations, these limiting factors will have to be considered: first, the greater susceptibility of the telephone to interfering noises, such as beat tones, and second the fact that the telephone requires two bands one for each direction of transmission and that these bands are required to be spaced a little apart in frequency. It is obvious that by controlling both types of channels from the same station they can be better co-ordinated in respect to frequency and general service use than if operated from separate stations, so that combined telephone-telegraph shore stations present interesting possibilities for the future.

Another method of operation, and one which requires fewer wave lengths, is that of superimposing the telephone and telegraph channel on the same carrier wave after the general manner of compositing long distance telephone lines with telegraph. This can be done by combining the two channels on one circuit as is done in wire practice and then modulating the combined channels upon the radio carrier. At the receiving end both channels can be detected simultaneously and then the channels separated by composite sets or filters. This method is mentioned to show the ultimate possibilities of combined operation and is not put forward as one which is sufficiently practical, all things considered, for use in the art of the immediate future.

Radio Linked with Transcontinental Line and Catalina Island

The ship-to-shore radio link was on several occasions connected with very long distance circuits in order to demonstrate the extreme conditions under which combined radio and wire operation are possible.

Perhaps the most interesting case is that in which the ship was linked up with the transcontinental telephone line and connected thru to Catalina Island in the Pacific thus bringing together the two oceans. The circuit arrangements for one of these demonstrations are given schematically in Fig. 24. Both the Deal Beach and Green

TRANSCONTINENTAL LINE WITH RADIO EXTENSIONS

Fig. 24

Harbor shore stations were used, since it was desired to reach the ship anywhere on her course from Boston to the Delaware capes. The demonstration was, therefore, also an example of connecting the ship into the land telephone system thru either of two shore stations. As a matter of fact, at one time the vessel could be reached thru both stations. It happened that the vessel was coming up the coast. The night before the demonstration the ship was communicated with thru the Deal Beach station and connected thru to Catalina Island for a rehearsal. For the demonstration of the following morning, connection was made thru Green Harbor. During both the rehearsal and the demonstration the operator on the vessel talked successfully, altho with some difficulty, with the Catalina Island operator, while New York listened in. This demonstration was

made for General J. J. Carty on February 14, 1921. An earlier demonstration of a similar nature, although not involving Green Harbor, was made for the delegates of the Preliminary International Communications Conference on October 21, 1920.

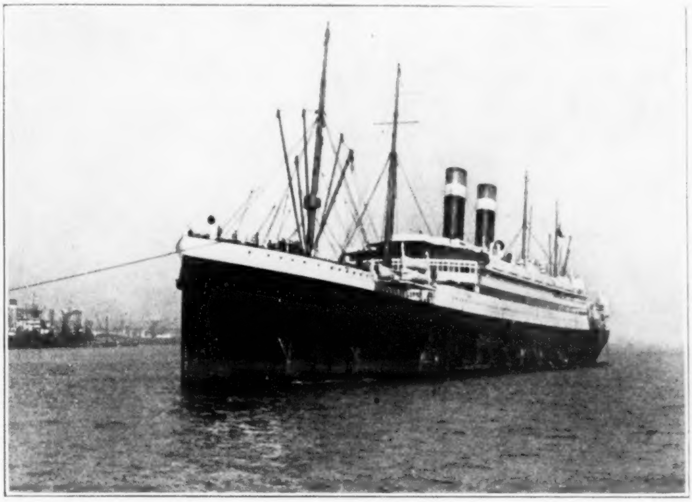


Fig. 25

Conclusions

The results of this development may be summed up as follows:

(1) It has realized a radio-telephone system capable of giving two-way transmission and meeting the requirements imposed by joint radio-wire operation.

(2) It has demonstrated the actual use of this radio-telephone system in a wire-radio toll circuit as a means for extending the telephone service of the country to include vessels at sea.

(3) The experiments have demonstrated also the practicability of multi-channel operation from a common land station whereby a number of land subscribers may be connected simultaneously to a number of different vessels.

(4) The transmission and operating tests show the difficulties attending the establishment and maintenance of the radio-telephone link to a moving vessel and the necessity for careful

adjustment of the transmission conditions of the circuit and for a diligent maintenance of these adjustments during operation.

(5) In the experiments in multi-channel operation and in simultaneous telephone and telegraph transmission from the same vessel, a beginning has been made in one of the most important problems concerned with the early application of radio telephony to the marine service, namely, that of the co-ordination between radio-telegraph and radio-telephone transmission. It is obvious that the general development of the art of selective transmission, as well as the entrance of radio telephony, calls for the use of purer carrier waves and of a minimum transmission band in radio telegraphy.

(6) As regards the important question of wave lengths, the development has shown that the relatively short waves employed in the experiments are satisfactory up to several hundred miles but that for longer distances longer wave lengths will be required. The difficulty of obtaining for the marine service a wave length sufficiently wide for permitting the handling of any considerable traffic is obvious. The band which can be allocated to this service will naturally be limited by the requirements of other services; and the intensiveness with which this band can be worked by closing up the frequency spacing between channels is limited by the consideration of intercommunication between different types of systems and by apparatus expense.

In general it may be said that the present development has contributed to the communication art the means whereby the universal land line telephone system may be extended to ships at sea. The actual giving of such service must await the working out of the economic problems involved and the necessary business and organization arrangements between the communication companies and the steamship companies.

DAILY RECORD SHEET

SHIP TO SHORE RADIO SERVICE

System Test

April 6, 1923.

TELEPHONE CALLS

		Destination	Total Time to Connect (hr. - min.)	Delayed	Interruptions	Cancelled	Completed Unsatisfactory	Completed Satisfactory	Remarks
From ship	1	N.Y.	41					✓	Good.
	2	N.Y.	37					✓	Good.
	3	N.Y.	16					✓	Good. Weak arc QRN.
	4	N.Y.	4					✓	Fair, (weak) bad arc QRM.
	5	Newark	3 - 53	P. (11.55 A.M. Cut till P.M. (2.14 P.M. Still out.				✓	Good, (weak [illegible], spk.)
	6	Long Branch	3 - 01	P. (11.53 A.M. Don't. ans. (2.13 P.M. Answ'd. Party out.				✓	Good, but weak.
	7	Asbury Park	2 - 40	P. 2.19 P.M. Out.				✓	Fair. [illegible] from [illegible].
	8	N.Y.	22					✓	Good.
	9	Newark	36	P. 3.00 P.M. Out of Bldg.				✓	Good, but weak.
	10	Asbury Park	-			2.47 P.M. Out till 5 P.M.			
	11	Newark	15					✓	Good. Quite noisy.
	12	N.Y.	50					✓	Good. Spk. QRM.
	13	Allenhurst, NJ	57					✓	Good. Weak QRN, 3 QRM.
	14	Asbury Park	29					✓	Good. (Bad spark QRM)
To ship	1	N.Y.C.	21					✓	Good.
	2	N.Y.C.	45					✓	Good. Ship has spark QRM.
	3	N.Y.C.	53					✓	Fair.
	4	N.Y.C.	47					✓	Fair.
	5	N.Y.C.	-			10.44 A.M; 3.43 P.M; don't ans.			
	6	N.Y.C.	19					✓	Good.
	7	N.Y.C.	11					✓	Good.
	8	N.Y.C.	12					✓	Good, Mr. Maxfield tries out transmitters.
	9	Spring Lake	18					✓	Good.
	10	N.Y.C.	20					✓	Good.
	11	N.Y.C.	03	P. (2.15 P.M. Not in (2.30 P.M. Till call)				✓	Good.
	12	N.Y.C.	3						Fair.
	13	N.Y.C.	-			3.02 P.M. Not there today.			
	14	N.Y.C.	26	P. 3.30 P.M. Line busy.				✓	Fair.
	15	N.Y.C.	18		E. 3.50-3.51 P.M. Cut off.			✓	Fair.
	16	N.Y.C.	9					✓	Fair.
	17	N.Y.C.	25					✓	Fair.

TELEGRAMS

		No. of Words	Total Time to Transmit	Actual Time on Traffic	Delay to Telephone Service
From ship	1				
To ship	1	114	13	8	0
	2	17	4	1	0

SUMMARY

Total calls	31
Completed	28
Cancelled	3
Long Distance	0
Toll	1
Local	30

Calls completed and satisfactory	28	90.3%
Calls completed and unsatisfactory	0	
Calls cancelled due to system failure	0	
Calls cancelled not due to system failure	3	9.7%
		100%

Average time

Radio A	13.6 mins.
" B	.16 "
Wire A	8.0 "
" B	0

$$\frac{\text{Total minutes operation}}{\text{Total calls}} = \frac{551}{31} = 17.7$$

The Contributors to this Issue

John R. Carson, B.S., Princeton, 1907; E.E., 1909; M.S., 1912; Research Department, Westinghouse Electric and Manufacturing Company; 1910–12; instructor of physics and electrical engineering, Princeton, 1912–14; American Telephone and Telegraph Company, Engineering Department, 1914–15; Patent Department, 1916–17; Engineering Department, 1918; Department of Development and Research, 1919—. Mr. Carson's work has been along theoretical lines and he has published several papers on theory of electric circuits and electric wave propagation.

Otto J. Zobel, A.B., Ripon College, 1909; A.M., Wisconsin, 1910; Ph.D., 1914; instructor in physics, 1910–15; instructor in physics, Minnesota, 1915–16; Engineering Department, American Telephone and Telegraph Company, 1916–19; Department of Development and Research, 1919—. Mr. Zobel has made important contributions to circuit theory in branches other than the subject of wave-filters.

J. N. Kirk, B.S., Purdue University, 1905; Engineering Department, New York Telephone Company, 1905–11; Plant Engineer for Texas, Southwestern Bell Telephone Company, 1912; Outside Plant Engineer, 1913–16; Engineering Department, American Telephone and Telegraph Company, 1917–19; Outside Plant Engineer, Department of Operation and Engineering, 1920—.

Helene C. Bateman, A.B., Barnard, 1917; Department of Operation and Engineering, American Telephone and Telegraph Company, 1917—. Mrs. Bateman has been engaged in research work relating to commercial engineering.

H. D. Arnold, Ph.B., Wesleyan, 1906; M.S., 1907; Ph.D., Chicago, 1911; assistant in physics, Wesleyan, 1906–07; Chicago, 1908; professor, Mt. Allison, 1909–10; Engineering Department of the Western Electric Company, Research Engineer, 1911—; Director of Research, 1923—. Dr. Arnold has been in direct charge of the development of the vacuum tube for telephone repeaters and radio purposes, and also other items of telephone equipment.

G. W. Elmen, B.S., University of Nebraska, 1902; M.A., 1904; Research Laboratories of the General Electric Company, 1904–06;

Engineering Department of the Western Electric Company, 1906—. One of Mr. Elmen's principal lines of work has been magnetic investigations.

Charles S. Demarest, E.E., University of Minnesota, 1911; American Telephone and Telegraph Company, Engineering Department, 1911–19; Department of Development and Research, 1919—. Mr. Demarest's work has been connected with the development of equipment for toll telephone systems, including that for long distance cables, carrier and radio systems.

H. W. Nichols, B.S., Armour Institute, 1908; E.E., 1911; M.S., Chicago, 1909; Ph.D., 1918; assistant professor electrical engineering, Armour, 1909–14; Engineering Department of Western Electric Company, 1914—. Since joining the research staff of the Western Electric Company, Dr. Nichols has been taking a leading part in the development of radio apparatus.

Lloyd Espenschied, Pratt Institute, 1909; United Wireless Telegraph Company as radio operator, summers, 1907–08; Telefunken Wireless Telegraph Company of America assistant engineer, 1909–10; American Telephone and Telegraph Company, Engineering Department and Department of Development and Research, 1910—. Took part in long distance radio telephone experiments from Washington to Hawaii and Paris, 1915; since then his work has been connected with the development of radio and carrier systems.

The Bell System Technical Journal

October, 1923

Mutual Impedances of Grounded Circuits

By GEORGE A. CAMPBELL

Synopsis: Formulas are derived for the direct-current mutual resistance and inductance between circuits grounded at the surface of the earth. For circuits composed of straight filaments, the mutual inductance is reduced to known Neumann integrals which involve only comparatively simple expressions for the case of horizontal, coplanar conductors above, below or on the surface of the earth. Numerical values for these integrals may be readily obtained from new and accurate graphs for straight filaments which meet at a point or start from a common perpendicular. It is shown that these new results supply a useful first approximation to the actual alternating-current mutual impedance of grounded circuits, when the frequency and extent of the circuits are not larger than occur in many practical applications.

1. Introduction

THE important discovery of the possibility of using the earth as the return conductor for electric telegraphic communication was announced by Steinheil in the *Comptes Rendus* of September 10, 1838, and throughout the entire development of telegraphy grounded circuits have been extensively employed. Considering the extensive application of such a capital discovery extending over a period of 85 years, it is surprising that so little is known quantitatively about grounded circuits. We have, however, long known that conditions are not of the extreme simplicity pictured under the early view that the earth acts as a reservoir presenting no resistance to the return current and introducing no interference between parallel returns. This view was expressed in 1857 by Bakewell as follows: "There is no mingling of currents, the electric current of each battery being kept as distinct as if separate wires were used both for the transmitted and the return currents. It would indeed be as impossible for the separate currents transmitted from the two batteries to be mingled together as it would be for the written contents of two letters enclosed in the same mail bag to intermix."

Measurements made a few years ago of the mutual impedances between grounded circuits which are restricted to a territory six miles square, at frequencies of 25 to 60 cycles per second, showed that within 10 per cent. the mutual reactance increased in the same ratio as the frequency. It was inferred that the effective inductance under the conditions of these tests was approximately the same as for direct current, or in other words, the incomplete penetration of the alternating currents into the earth was not of controlling importance in tests upon this scale.

This led to my making a theoretical investigation of the mutual inductance between direct-current grounded circuits which did, in fact, show that the calculated numerical results are in reasonable agreement with these actual experimental data. It is the purpose of this paper to describe this work; the mathematical discussion of the theoretical corrections for the incomplete penetration of alternating currents into the earth will form the subject of another paper.

2. Distribution of Current, Potential and Magnetic Force with Direct-Current Earth Return Flow

On the assumption of an infinite earth of uniform resistivity the lines of flow and the equipotential surfaces for a direct current I entering the earth at a point source at A and leaving the earth at a point sink at B, both A and B being on the surface of the earth, assumed flat, and the distance $AB=2b$, are given by the equations,[1]

$$\left.\begin{aligned}\frac{C}{I} &= \tfrac{1}{2}(\cos\theta_1-\cos\theta_2)\\ &= \tfrac{1}{2}\left(\frac{x_1}{r_1}-\frac{x_2}{r_2}\right)\\ &= \sin\tfrac{1}{2}(\theta_1+\theta_2)\sin\tfrac{1}{2}(\theta_2-\theta_1)\\ &= b\,\frac{\sin^2\theta}{r},\ \text{if } \frac{b}{r} \text{ is small,}\end{aligned}\right\} \qquad (1)$$

$$\left.\begin{aligned}\frac{V}{I} &= \frac{\rho}{2\pi}\left(\frac{1}{r_1}-\frac{1}{r_2}\right)\\ &= -\frac{b\rho}{\pi}\,\frac{\cos\theta}{r^2},\ \text{if } \frac{b}{r} \text{ is small,}\end{aligned}\right\} \qquad (2)$$

where C is the total current flowing in the earth, from the source at A to the sink at B, outside the current sheet of revolution defined by (1); and V is the potential, with respect to the midplane, upon the equipotential surface of revolution defined by (2).

These equipotential lines and stream lines are identical with the equipotential lines and lines of force for a uniformly magnetized filament. The formulas may be checked by regarding the return flow from A to B as being due to the superposition of two flows, a

[1] The coordinates used in this paper (x_1, y, z), (x_2, y, z), (x, y, z) and (r_1, θ_1, ϕ), (r_2, θ_2, ϕ), (r, θ, ϕ) are rectangular and spherical coordinates with origins at A, B and the midpoint of AB, respectively, the direction AB being the polar axis or positive x-axis in all cases, z being vertical, and ϕ being measured from the earth's surface in the plane perpendicular to AB.

return flow of direct current I from the point A to some infinitely distant point and a second return flow of direct current I from this infinitely distant point to the point B. For these component flows the current diverges from A or converges towards B radially and with equal intensity in all directions in the earth; the total current for one of the component flows flowing through any surface in the earth will thus be equal to $I/2\pi$ times the solid angle subtended at A or B, respectively, by the boundary of the surface, since the entire solid angle filled by the earth at a point on the surface is 2π. The total radial flow from A through the lower half of the circular cone having its axis in AB, the elements of the cone making the angle θ_1 with AB, is $\frac{1}{2}I(1-\cos\ \theta_1)$; similarly, the total radial flow toward B through the lower half of a cone with the angle $\pi-\theta_2$ will be $\frac{1}{2}I(1+\cos\theta_2)$. For the combined superposed flows the total current flowing through the semicircle in which the cones intersect is the sum of these two values or $\frac{1}{2}I(2-\cos\ \theta_1+\cos\ \theta_2)$, from which (1) is immediately obtained, since the total current flowing in the earth from A to B is I. This assumes that the semicircle lies between A and B, but the same formula holds for the entire current sheet of revolution. The lines of flow in the earth are symmetrical about AB and lie in planes through AB, since, in the earth, both component flows are symmetrical about AB.

For the component flows the equipotential surfaces are hemispherical and, since the resistance of a hemispherical shell of radius r, thickness dr, is $\rho dr/2\pi r^2$, the potentials at distances r_1 or r_2 from A or B, referred to the potential at infinity, are $I\rho/2\pi r_1$ or $-I\rho/2\pi r_2$, respectively, from which equation (2) follows by addition.

Fig. 1 accurately reproduces the flow and equipotential lines as given by formulas (1) and (2). At the midpoint of a line of flow its distance from each electrode is $r_1=r_2=bI/C$ and it may be shown that every other point of a line of flow is at a still shorter distance from the nearer electrode. It follows, for example, that less than 1/10 of the total current reaches, in its flow through the earth, any point lying at a distance greater than $5AB$ from the line AB connecting the electrodes.

If a uniform radial flow of current I in the horizon plane converging on the point A is combined with the uniform radial flow in the earth outward from A, we have a closed flow which is symmetrical about the vertical axis through A. Below the horizon plane the magnetic lines of force will be horizontal circles and the magnetic force at any point distant r_1 from A, α being the angle included between r_1 and the nadir, is $H=2I(1-\cos\alpha)/(r_1\sin\alpha)$

$=2I\ r_1^{-1} \tan (\alpha/2)$. Above the horizon plane there is no magnetic field, since any magnetic lines of force are, by symmetry, horizontal circles and the intensity is zero, since there is no current threading any horizontal circle above the surface of the earth.

Superposing this closed flow and a similar closed flow through B from the earth to the horizon plane, we obtain a closed flow from A

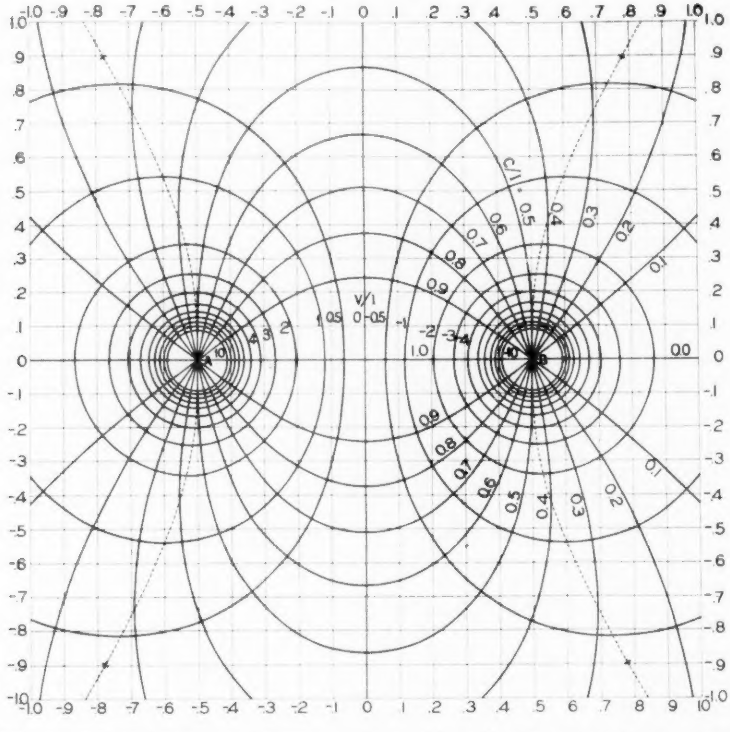


Fig. 1—Flow and equipotential lines on the earth's surface for an earth return flow from A to B. C/I is the fraction of the current flowing in the earth outside the flow surface of revolution; V/I is the resistance to the flow of the portion of the earth lying between the equipotential surface of revolution and the mid or zero potential plane, if the earth's resistivity is $\rho = 2\pi$. The flow and equipotential lines through each point on the dotted curve are perpendicular and parallel, respectively, to AB

to B in the earth and back from B to A in the horizon plane. The magnetic field for this closed flow, being the sum of the magnetic fields for the component flows, will be zero above the horizon plane, while below it will consist of lines of force in horizontal planes. This

result also applies to any closed flow which does not extend above the horizon plane and may be resolved into any number of component flows, each of which is radially symmetrical about a vertical axis.

3. Mutual Resistance of Grounded Circuits

By definition, if e.m.f.'s E and e in grounded conductors AB and ab produce the currents I and $i=0$ in the conductors, the mutual impedance between the two conductors is e/I. In the present case we are dealing with direct current and thus the mutual impedance is a mutual resistance Q, and by (2) its value is [2]

$$\begin{aligned} Q &= \frac{\rho}{2\pi}\left(\frac{1}{Aa}-\frac{1}{Ab}-\frac{1}{Ba}+\frac{1}{Bb}\right) \\ &= \frac{\rho}{2\pi}\int\int\frac{d^2}{dSds}\left(\frac{1}{R}\right)dSds \\ &= \frac{\rho}{2\pi}\int\int\frac{-2dUdu+dVdv+dWdw}{R^3} \\ &= \frac{\rho}{2\pi}\int\left\{\frac{\cos(\theta_1-\epsilon)}{r_1^2}-\frac{\cos(\theta_2-\epsilon)}{r_2^2}\right\}ds. \end{aligned} \qquad (3)$$

The third form of (3) shows that the mutual resistance falls off as the inverse third power of the distance between grounded circuits when this distance has become large compared with the length of these circuits between grounding points.

The first form of (3) shows that the mutual resistance between grounded circuits does not depend upon the location of the conductors but only upon the location of the terminal grounding points A, B, a, b.

The mutual resistance for the case $\rho=2\pi$ is obtained from Fig. 1 by taking the value of V/I at the point corresponding to a reduced by its value at the point corresponding to b; if b is anywhere on the center line, for which $V/I=0$, the diagram gives directly the value of the mutual resistance. Employing ordinary units the diagram gives the mutual resistance directly in ohms if $AB=1$ mile and the earth has a resistivity of about one million ohms per centimeter cube (more exactly 1.011×10^6) which is its actual order of magnitude.

[2] In addition to the earlier notation there are employed in the different expressions for formula (3), and also in formula (5) below, the following: R is the distance between two elements dS and ds of any two paths extending from A to B and a to b; the rectangular projections of these elements along and perpendicular to R are dU, dV, dW and du, dv, dw, the two sets being parallel and with the same positive directions; θ_1, θ_2, ϵ are the angles which r_1, r_2 and ds make with AB, when the path ab lies in a plane with A and B.

4. Neumann Integrals for Return Flows

The required mutual inductances of grounded circuits will be found by means of the Neumann integral

$$N=\int\int\int\int(\cos\epsilon/r)dI\,dS\,di\,ds$$

extended over every current filament in both flows. Since the earth return portions of the two flows are independent of the flows in the arbitrarily located conductors on the earth's surface, it is convenient to divide the Neumann integral into four partial integrals which involve either no return flow, one return flow or both return flows according to the following formula [3]

$$\begin{aligned} N_{(\mathcal{X}-\mathcal{E})(x-e)} &= N_{\mathcal{X}x}-N_{\mathcal{X}e}-N_{\mathcal{E}x}+N_{\mathcal{E}e} \\ &= N_{\mathcal{X}x}-(\tfrac{1}{2}+\tfrac{1}{2}-1)\Delta, \text{ by Table I,} \\ &= N_{\mathcal{X}x}. \end{aligned} \tag{4}$$

Checking the entries of Table I may be accomplished without performing more than two integrations. It will be convenient to make the integrals somewhat more general than is required in checking the table and find $N_{\mathcal{F}s}$ and $N_{\mathcal{X}'s}$ where $\mathcal{F}$ is any flow in space from A to B, which need not be coplanar points with the terminals a and b of s, and $\mathcal{X}'$ is any flow in a plane parallel to the horizon plane between terminal points A' and B'.

Consider first the part of a space return flow s which is radial from a in connection with an element dS on any filament of current dI of a flow $\mathcal{F}$ from A to B. The component dx of dS along the line x from a to ds is the only component which need be considered, since by symmetry the normal component contributes nothing to the Neumann integral. As the total radial flow is to be taken equal to unity, the amount flowing out through a ring, taken as the volume element, lying between the spheres of radii s and $s+ds$ and between the cones making angles θ and $\theta+d\theta$ with x will be $\frac{1}{2}\sin\theta\,d\theta$. If this ring lies at a distance r from dS the Neumann integral will be

$$\begin{aligned} N &= \int_{Aa}^{Ba}dx\int_0^\infty ds\int_0^\pi \frac{\cos\theta\sin\theta\,d\theta}{2r},\ r^2=x^2+s^2-2xs\cos\theta, \\ &= \frac{1}{4}\int_{Aa}^{Ba}\frac{dx}{x^2}\int_0^\infty\frac{ds}{s^2}\int_{|x-s|}^{(x+s)}(x^2+s^2-r^2)dr \end{aligned}$$

[3] Each term indicates the Neumann integral for the pair of flows designated by the script letter subscripts, as explained in the note accompanying Table I. Both $(\mathcal{X}-\mathcal{E})$ and $(x-e)$ are arbitrary flows on the earth's surface closed by earth return flows from A to B and from a to b, respectively.

$$= \frac{1}{3}\int_{Aa}^{Ba} \frac{dx}{x^2}\left\{\int_0^x s\,ds + x^3\int_x^\infty \frac{ds}{s^2}\right\}$$

$$= \frac{1}{2}\int_{Aa}^{Ba} dx = \tfrac{1}{2}(Ba - Aa).$$

TABLE I

Entries are the value of k in the formula $N = k\Delta$ for the Neumann integral between the specified flows, $\Delta = -Aa + Ab + Ba - Bb$, or $2AB$ if $A = a$, $B = b$, and points A B, a, b are all on the earth's surface.

Flows†	h	s	e	a	n	z	$(h-s)$	$(h-e)$	$(h-a)$	$(h-n)$	$(h-z)$	$(e-a)$
Any surface $=\mathscr{X}$	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	1	0
Space return $=\mathscr{S}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0	0	0
Earth return $=\mathscr{E}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0	$\frac{3}{2}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$	$\frac{1}{2}$	-1	1	1
Air return $=\mathscr{A}$	$\frac{1}{2}$	$\frac{1}{2}$	0	1	$-\frac{1}{2}$	$\frac{3}{2}$	0	$\frac{1}{2}$	$-\frac{1}{2}$	1	-1	-1
Nadir return $=\mathscr{N}$	0	$\frac{1}{2}$	$\frac{3}{2}$	$-\frac{1}{2}$	*	-1	$-\frac{1}{2}$	$-\frac{3}{2}$	$\frac{1}{2}$	*	1	2
Zenith return $=\mathscr{Z}$	0	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$	-1	*	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{3}{2}$	1	*	-2
Closed $(\mathscr{X}-\mathscr{S})$	$\frac{1}{2}$	0	0	0	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	1	0
" $(\mathscr{X}-\mathscr{E})$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{3}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0	2	0	-1
" $(\mathscr{X}-\mathscr{A})$	$\frac{1}{2}$	0	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{3}{2}$	$\frac{1}{2}$	0	1	0	2	1
" $(\mathscr{X}-\mathscr{N})$	1	0	-1	1	*	1	1	2	0	*	0	-2
" $(\mathscr{X}-\mathscr{Z})$	1	0	1	-1	1	*	1	0	2	0	*	2
" $(\mathscr{E}-\mathscr{A})$	0	0	1	-1	2	-2	0	-1	1	-2	2	2

* Not of the assumed form and in general infinite.

† Each type of flow is designated by a script letter. Any flow in the surface of the earth, assumed to be a plane, is designated by $\mathscr{X}$, there being no restriction on the number of filaments of current but each filament must start from a common point, A, and extends to another common point, B. A special case of $\mathscr{X}$ is $\mathscr{H}$ the horizon return flow made up of two superposed uniform radial flows, from A to every point of the horizon and back to B. The space return flow, $\mathscr{S}$, is made up of two superposed uniform radial flows, one outward in all directions from the point A and the other inward from all directions toward the point B. The earth return and air return flows, $\mathscr{E}$ and $\mathscr{A}$, are similar except that the flows are uniformly distributed over all directions in the earth and air, respectively. The nadir return and zenith return flows, $\mathscr{N}$ and $\mathscr{Z}$, consist of two infinite vertical filaments from A and back to B by way of the nadir and zenith, respectively. Small script letters indicate similar types of flow with any independent terminal points, a and b. Differences such as $(\mathscr{X}-\mathscr{S})$ designate closed flows; thus $(\mathscr{X}-\mathscr{E})$ designates any flow on the earth's surface from A to B where it enters the earth and after spreading out uniformly through the earth returns to the terminal A, thus closing the flow.

To this is to be added the corresponding expression $\frac{1}{2}(Ab-Bb)$ for the radial flow converging on b, giving the result $\frac{1}{2}\Delta$. As the path of the line of flow between A and B does not enter into the result, it is immaterial whether the flow is confined to a single filament or is spread out in any way whatsoever in space, provided only all stream lines extend from A to B, as assumed for $\mathcal{F}$. Thus

$$N_{\mathcal{F}_s} = \tfrac{1}{2}\Delta = N_{\mathcal{S}_f}.$$

To find $N_{\mathcal{X}'A}$ let r and s be the distances of any element dS of a line of flow forming a part of $\mathcal{X}'$ from a and from any element of a plane radial flow from a, respectively, the projections of r and s on either of the planes being x and y, which include the angle ϕ; $s^2=y^2+z^2$, $r^2=x^2+z^2$. The component of dS parallel to x will be dx and this is the only component which need be considered, since, on account of the symmetry of the radial flow, the normal component in the plane of flow contributes nothing to the integral.

$$\begin{aligned}
N &= \int_{Aa}^{Ba} dx \int_0^\infty \int_0^{2\pi} \frac{x-y\cos\phi}{x^2+y^2-2xy\cos\phi}\,\frac{y\,dy\,d\phi}{2\pi s} \\
&= \frac{1}{2\pi}\int_{Aa}^{Ba} \frac{dx}{x}\int_z^\infty ds \int_0^\pi \left(1+\frac{x^2-y^2}{x^2+y^2-2xy\cos\phi}\right)d\phi \\
&= \frac{1}{2\pi}\int_{Aa}^{Ba} \frac{dx}{x}\int_z^\infty ds\left[\phi-\sin^{-1}\frac{(x^2+y^2)\cos\phi-2xy}{x^2+y^2-2xy\cos\phi}\right]_0^\pi \\
&= \int_{Aa}^{Ba}\frac{dx}{x}\int_z^r ds,
\end{aligned}$$

since inspection shows that the two values of the definite integral 2π and 0 are to be used for $s \lesseqgtr r$ respectively, and therefore

$$\begin{aligned}
N &= \int_{Aa}^{Ba} \frac{r-z}{x}\,dx \\
&= \int_{A'a}^{B'a} \frac{r\,dr}{r+z} \\
&= \Big[r - z\log(r+z)\Big]_{A'a}^{B'a} \\
&= (B'a-A'a) - z\log\frac{B'a+z}{A'a+z}.
\end{aligned}$$

This is for the radial flow from a. Adding the corresponding expression

for the radial flow towards b, we have finally for the complete integral

$$N\mathscr{X}'_{h}=(-A'a+A'b+B'a-B'b)-z\log\frac{(A'b+z)\,(B'a+z)}{(A'a+z)\,(B'b+z)}, \quad \textbf{(4a)}$$

which becomes, if $z=0$,

$$N\mathscr{X}_{h}=\Delta=N\mathscr{H}_{x}.$$

The first line of Table I can now be filled in at once since the integrations have shown that the first two values of k are 1 and $\frac{1}{2}$; the next two entries are also $\frac{1}{2}$ since by symmetry $N\mathscr{X}_{s}=N\mathscr{X}_{e}=N\mathscr{X}_{a}$; $N\mathscr{X}_{n}=N\mathscr{X}_{z}=0$ since the nadir and zenith flows are perpendicular to the $\mathscr{X}$ flow in the horizon plane. The remaining six entries in the first row are for closed flows which are expressed as differences between the flows already considered, and the corresponding k's are the differences of the k's for the component flows. The first column of k's may also be filled in, the table being symmetrical, since interchanging capital and small script letters leaves N unchanged and h is a special case of x.

The second row of the table involves only special cases of $N\mathscr{S}_{f}$ and only the values $\frac{1}{2}$ and 0 occur.

From the flows included in the table nine pairs of closed flows may be formed having zero mutual inductances, because one of the closed flows of each pair has no magnetic field below the surface of the earth and the other closed flow includes no current above the surface of the earth, and thus there is no interlinkage of induction between the two closed flows. The portion of Table I referred to is repeated in Table II, where the flows at the top are those for which any difference such as $(h-a)$ has no magnetic field below the earth's surface, just as $(\mathscr{H}-\mathscr{E})$ has no magnetic field above the earth's surface, as was proved above, while no flow at the side penetrates above the earth's surface.

TABLE II

	h	a	z
$\mathscr{X}$	1	$\frac{1}{2}$	0
$\mathscr{E}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
$\mathscr{N}$	0	$-\frac{1}{2}$	-1

The top row of Table II includes only values of k already found, and the remainder of the first column follows from symmetry and the

fact that $\mathcal{h}$ is a special case of the general flow $\mathcal{x}$. Any other entry in Table II is now found in terms of three of the entries in this border, thus

$$0 = N(\mathcal{X} - \mathcal{E})(\mathcal{h} - \mathcal{a}) = N\mathcal{X}\mathcal{h} - N\mathcal{X}\mathcal{a} - N\mathcal{E}\mathcal{h} + N\mathcal{E}\mathcal{a} = (1 - \tfrac{1}{2} - \tfrac{1}{2})\Delta + N\mathcal{E}\mathcal{a},$$

and we have the interesting result $N\mathcal{E}\mathcal{a} = 0$; the remainder of the table follows.

The result $N\eta_z = -\Delta$ may be readily checked directly since it involves only the mutual Neumann integral between straight parallel filaments, and by using the expanded form of the integral for equal filaments beginning at a common perpendicular with opposite positive directions[4] the result can be written down at once.

The important difference Δ which is utilized in Table I may be expressed in the following useful forms:

$$\left.\begin{aligned} \Delta &= (-Aa + Ab + Ba - Bb) \\ &= -\int\int \frac{d^2R}{dSds}\, dSds \\ &= \int\int \frac{dVdv + dWdw}{R} \\ &= 2\int \sin \tfrac{1}{2}(\theta_2 - \theta_1) \sin \left[\tfrac{1}{2}(\theta_1 + \theta_2) - \epsilon\right] ds, \end{aligned}\right\} \tag{5}$$

where the notation is the same as for formula (3) above. The third form of (5) shows that when the separation R is great the mutual inductance varies inversely as the first power of the separation.

5. Mutual Inductance Between Grounded Circuits Lying on the Surface of the Earth

It has now been shown that *for direct currents the mutual inductance between grounded circuits consisting of conductors lying on the surface of the earth and grounded at their terminals is equal to the mutual Neumann integral between the conductors alone*, since in the complete Neumann integral for the closed flows the total contribution of those parts which involve the ground returns is zero. For low frequencies the effective inductance can differ but little from the direct-current inductance, and it is therefore of practical importance to investigate

[4] Mutual Inductances of Circuits Composed of Straight Wires. *Physical Review*, 5, pp. 452–458, June, 1915, formula (6).

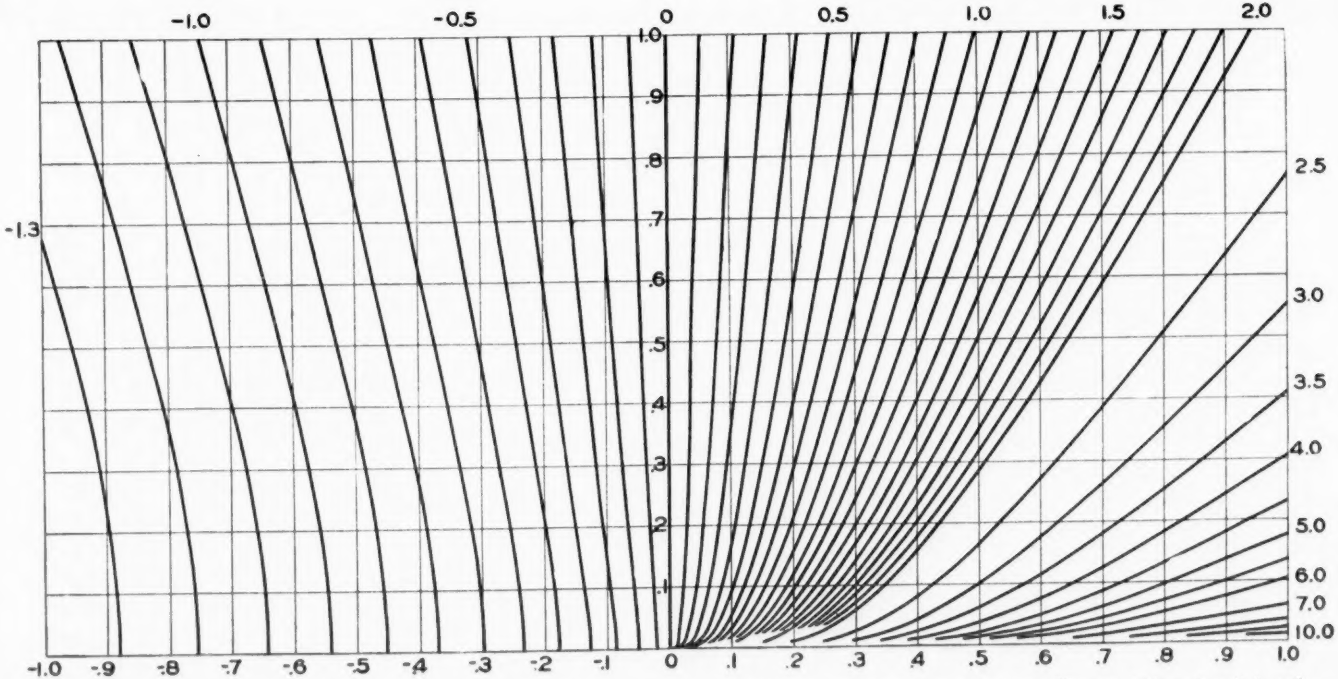


Fig. 2—Contour lines of the mutual Neumann integral between two straight filaments meeting at a point, one filament being of unit length

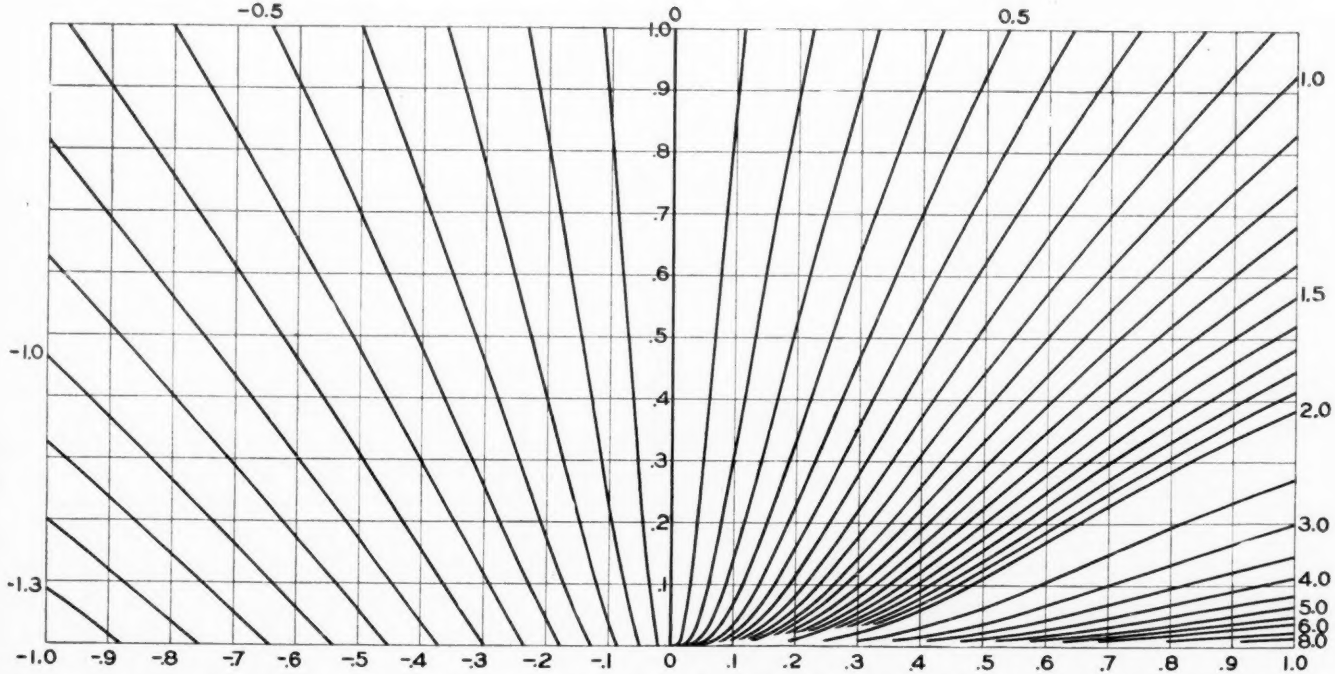


Fig. 3—Contour lines of the mutual Neumann integral between two parallel filaments starting at a common perpendicular, one filament being of unit length

the numerical magnitude of these Neumann integrals between grounded circuits. In order to visualize the magnitudes involved and supply means by which they may be readily calculated, a number of diagrams have been prepared for the important case of straight conductors.[5]

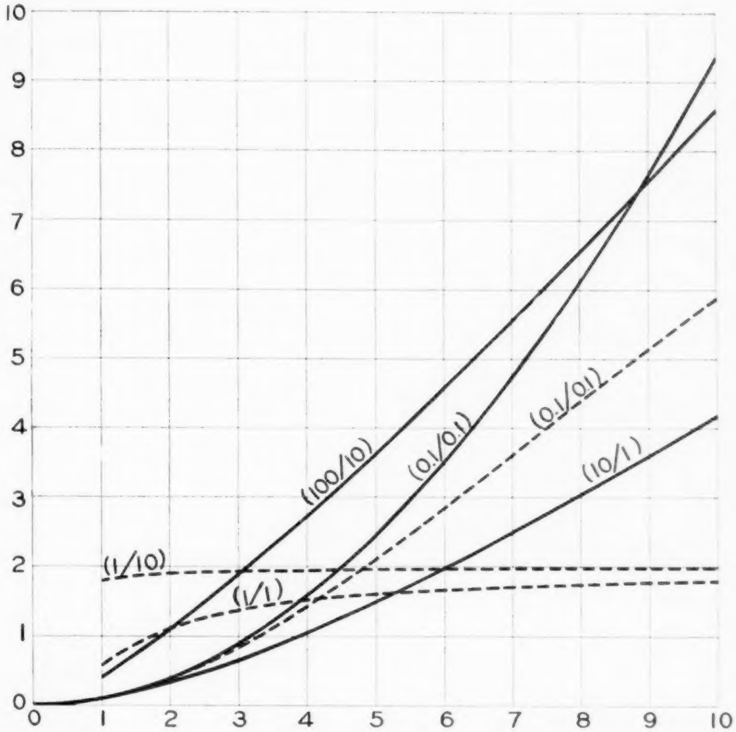


Fig. 4—Mutual Neumann integral between filaments forming opposite sides of a rectangle with unit separation. The dashed curves show the mutual resistance between the filaments if the circuits are grounded through the earth and its resistivity is $\rho = 2\pi$. The numerator and denominator of the bracketed fraction on each section of the curve show the factors by which the vertical and horizontal scales must be multiplied for use on this section

If the two straight conductors OA, Oa start from a common point O, the mutual Neumann integral is shown by Fig. 2; the curves give the locus of terminal a for constant values of the integral when the other conductor OA is the unit base. The Neumann integral between

[5] The necessary formulas are given in the paper loc. cit. Additional transformations of these formulas are given in the appendix to the present paper.

any two straight filaments $\mathcal{R}$ and r having terminals A, B and a, b may be expressed as

$$N_{\mathcal{R}r} = N_{(Aa)} - N_{(Ab)} - N_{(Ba)} + N_{(Bb)}, \tag{6}$$

where $N_{(Aa)}$ stands for the Neumann integral between the two straight filaments, OA and Oa, beginning at O, the point of intersection of $\mathcal{R}$ and r, extended if necessary, and ending at the terminals A and a; thus

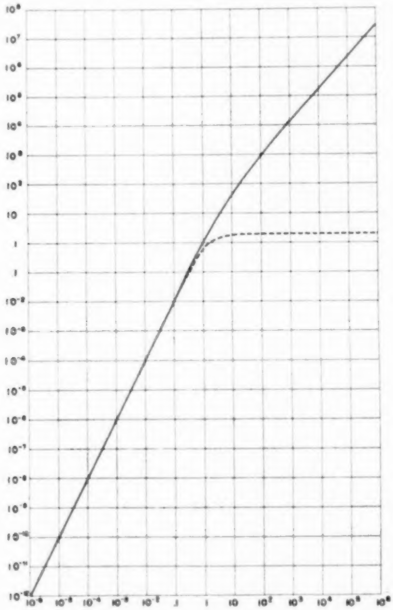


Fig. 5—Mutual Neumann integral between filaments forming opposite sides of a rectangle with unit separation. The dashed curves show the mutual resistance between the filaments if the circuits are grounded through the earth and its resistivity is $\rho = 2\pi$. This is Fig. 4, but with logarithmic scales

the integral between any two filaments which would intersect within a finite distance may be readily found after reading four values from Fig. 2.

In the special case of parallel filaments Fig. 2 fails; the corresponding curves for this case are presented by Fig. 3, which assumes a unit base filament and a parallel filament starting at a point on the left hand perpendicular to the base. Differences will give the general case

of parallel filaments which do not start at a common perpendicular, which may thus be derived from Fig. 3.

The mutual Neumann integral between any two parallel filaments may also be obtained by means of the formula

$$N_{\mathcal{R}r} = \tfrac{1}{2}[-N_{(Aa)} + N_{(Ab)} + N_{(Ba)} - N_{(Bb)}], \tag{7}$$

where $N_{(Aa)}$ now stands for the Neumann integral between the projections of Aa on $\mathcal{R}$ and r, extended if necessary, the projections

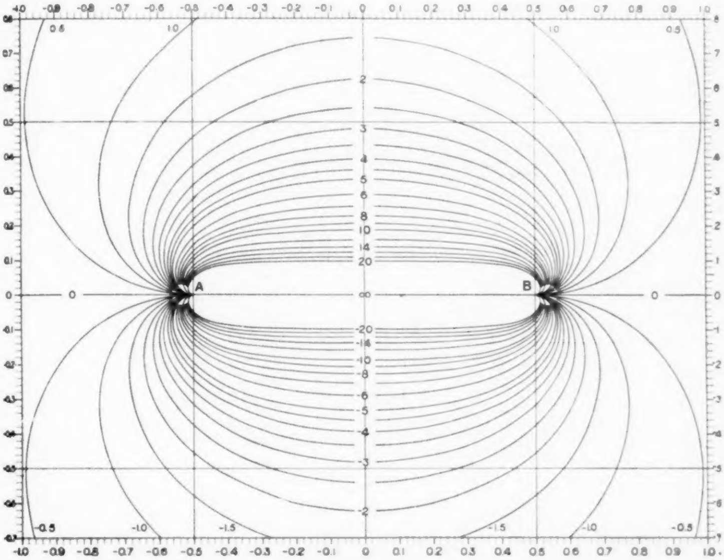


Fig. 6—Contour lines of the mutual Neumann integral between a counter-clockwise small loop on the surface of the earth, per unit area, and a straight grounded filament AB of unit length

having the same or opposite positive directions in agreement with $\mathcal{R}$ and r. This formula for the mutual Neumann integral presents the advantage of requiring only a single entry diagram, which is supplied by Fig. 4 and on a logarithmic scale by Fig. 5.

The mutual inductance may be required between a small, closed loop lying upon, but insulated from, the surface of the earth and a straight grounded conductor. The value depends upon the location, area and assumed positive direction around the loop, but is independent of the shape of the loop. Contour curves for the mutual inductance per unit area of the loop are given by Figs. 6 and 7; the

positive direction around the loop is counter-clockwise; the straight grounded conductor AB of Fig. 6 is of unit length while in Fig. 7 grounded terminal B alone appears, terminal A being at an infinite

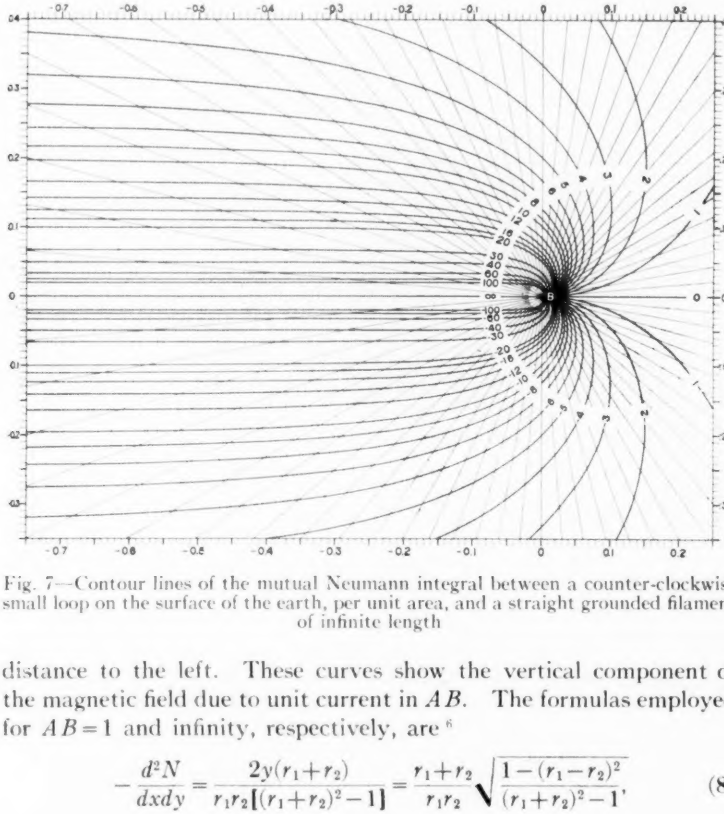


Fig. 7—Contour lines of the mutual Neumann integral between a counter-clockwise small loop on the surface of the earth, per unit area, and a straight grounded filament of infinite length

distance to the left. These curves show the vertical component of the magnetic field due to unit current in AB. The formulas employed for $AB=1$ and infinity, respectively, are [6]

$$-\frac{d^2N}{dxdy}=\frac{2y(r_1+r_2)}{r_1r_2[(r_1+r_2)^2-1]}=\frac{r_1+r_2}{r_1r_2}\sqrt{\frac{1-(r_1-r_2)^2}{(r_1+r_2)^2-1}}, \tag{8}$$

$$-\frac{d^2N}{dxdy}=\frac{y}{r_2(x_2+r_2)}=\frac{1}{r_2}\tan\tfrac{1}{2}\theta_2. \tag{9}$$

Large loop mutual inductances may be calculated either by integrating the value of $-d^2N/dxdy$ over the loop or by integrating the value of dN/ds around the boundary. If the boundary may be approximated to by a broken straight line, the curves of Figs. 2 and 3 may be employed.

[6] These formulas may be derived by differentiating (8) of the paper loc. cit., $dN/dx=2\tanh^{-1}[AB/(A_s+_sB)]$, with respect to y.

6. Mutual Impedances for Conductors Lying on the Surface of the Earth

In order to arrive as directly as possible at a concrete numerical idea of the magnitudes and angles occurring in the mutual impedances encountered in engineering work, we may advantageously start with the following specific constants:

Base Length (AB, OA or Aa) . . . . 1 Mile,

Frequency of the Alternating Current . . . . . 1 Kilocycle,

Resistivity of the Earth per Centimeter Cube . 1 Megohm,

which are of the right order of magnitude and make the factors

$$\rho/(2\pi\, AB) = 10^6/(2\pi\ 0.1609 \times 10^6) = 0.989,$$

$$2\pi f \times AB \times 10^{-9} = 2\pi\ 0.1609 = 1.011,$$

which are both equal to unity within about 1 per cent., so that the approximate resistance and reactance components of mutual impedances may be read directly from Figs. 1–8 without applying multipliers. On the other hand, when mutual impedances are required for other lengths, frequencies and specific resistances, the correcting factors are readily applied. The tangent of the angle of the mutual impedance is proportional to the frequency, to the square of the linear dimensions of the circuits and to the reciprocal of the earth's resistivity.

Grounded circuits separated by a distance large compared with the dimensions of the circuits have a mutual impedance with negligible resistance component since ultimately this component varies inversely as the cube of the distance by (3), whereas ultimately the mutual reactance varies inversely with the distance.

For two parallel grounded conductors separated by one mile and forming opposite sides of a rectangle, the two components of the mutual impedance are shown by Fig. 5 for the assumed constants (approximately a kilocycle and a megohm). The resistance component of the mutual impedance is then always less than the reactance component; when the rectangle becomes a square, the mutual impedance angle is $\tan^{-1}1.595 = 57.9°$. Reducing the frequency to 0.627 kilocycles or reducing the side of the square to 0.792 miles or increasing the resistivity to 1.595 megohms would reduce this angle to 45°.

Consider two straight grounded conductors AB and ab, the latter distance being small compared with the other dimensions of the

system and assume that while ground a is fixed, ground b is rotated about a in a circle of fixed radius ab. This will vary the mutual impedance between the two grounded conductors. The reactance component will vary as the cosine of the angle between ab and AB. The resistance component will also vary as the cosine of an angle, measured, in general, from a direction other than AB. The maximum resistance component will also differ, in general, from the maximum reactance component. The locus of the mutual impedance obtained for all positions of ab will be an ellipse. The ellipse becomes a straight line when ground a lies on the bisecting normal of AB, for then the direction giving the maximum resistance component is parallel to AB and thus the same as for the maximum reactance component. The straight line limit is also obtained when ground a lies on the prolongation of AB in either direction, for the resistance component then has its maximum value in the direction opposite to AB. The maximum resistance component will be perpendicular to AB at the points on Fig. 1 where the C/I contour, if drawn, would be vertical. The locus of these points is

$$y^2=(x^2-b^2)^{2/3}[(x+b)^{2/3}+(x-b)^{2/3}]. \tag{10}$$

At remote points on this locus $y=\pm\sqrt{2}\,x$ and the maximum resistance is negligible compared with the maximum reactance since the circuits are widely separated, while in the neighborhood of A and B it is the maximum reactance which is negligible compared with the maximum resistance. At some intermediate point the two maxima are equal, and, since they occur for directions differing by 90°, the elliptical impedance locus becomes a circle; and the mutual impedance between the two grounded circuits does not change in magnitude as ground b is rotated about ground a. For the assumed constants (1 mile, 1 kilocycle and 1 megohm) this point lies at distances of 1.562 and 0.939 miles from the two terminals A and B, and its four possible locations are shown by the four small crosses on Fig. 1.

If a is rotated counter-clockwise about AB, the direction giving the maximum resistance component also rotates counter-clockwise, making two complete revolutions, while the ground a makes one revolution about AB.

7. Equivalent Ground Plane

To a first approximation the direct-current mutual inductance between two straight conductors AB and ab, forming opposite sides of a rectangle on the surface of the earth, at a separation Aa which

is small compared with the length AB is, by (25) of the appendix, neglecting the first and higher powers of $1/s$,[7]

$$N=2AB\log\frac{\frac{2}{e}AB}{Aa}=2AB\log\frac{0.736\,AB}{Aa}. \tag{11}$$

This expression has the form $(2\,l\log s/r)$ of the commonly employed mutual inductance formula for two long parallel conductors, each of length l, separated by distance r, the common return being a perfectly conducting earth in which the image of each conductor is at the distance s from the other physical conductor. For our direct-current case, therefore, the effective distance to the images is about $\frac{3}{4}$ of the length of either grounded conductor. Since this distance is by assumption large compared with the distance between the conductors, the images are approximately at this same depth below the actual surface of the earth, and the hypothetical perfectly conducting earth would be at one-half this depth, or $\frac{3}{8}AB$. The effective image distance is necessarily directly proportional to the dimensions of the grounded circuits and independent of the earth's resistivity because the shape and relative distribution of the lines of flow are independent of the resistivity and of the length of the grounded circuits. Inspection of Fig. 1 shows that somewhat over $\frac{1}{2}$ of the return flow attains a distance $\frac{3}{4}AB$ from AB, while the remainder of the current remains closer to the grounded conductor.

It may be inquired what would be the effect of confining both return currents to a thin uniform conducting layer on the earth's surface, so that they become horizon return flows. For the closed flows $(\mathscr{X}-\mathscr{H})$ and $(\boldsymbol{x}-\boldsymbol{h})$ in general and the particular flows $(\mathscr{R}-\mathscr{H})$ and $(\boldsymbol{r}-\boldsymbol{h})$, where $\mathscr{R}$ and $\boldsymbol{r}$ are close parallel straight conductors,

$$\begin{aligned}
N_{(\mathscr{X}-\mathscr{H})(\boldsymbol{x}-\boldsymbol{h})} &= N_{\mathscr{X}\boldsymbol{x}}-N_{\mathscr{X}\boldsymbol{h}}-N_{\mathscr{H}\boldsymbol{x}}+N_{\mathscr{H}\boldsymbol{h}}\\
&= N_{\mathscr{X}\boldsymbol{x}}-\Delta;\\
N_{(\mathscr{R}-\mathscr{H})(\boldsymbol{r}-\boldsymbol{h})} &= N_{\mathscr{R}\boldsymbol{r}}-2Ab+2Aa\\
&= 2AB\log\frac{\frac{2}{e^2}AB}{Aa}=2AB\log\frac{0.271AB}{Aa}.
\end{aligned} \tag{12}$$

[7] If the term $1/s=Aa/AB$ of the expansion is retained the equivalent ground plane has the depth $(AB+Aa)/e$ and thus becomes deeper as ab is moved away from AB. But the equivalent ground plane may be kept fixed at the distance AB/e from AB provided it is tipped at the angle $\sin^{-1}2/e=47°$ so that ab moves away from the ground plane as it moves away from AB. If it were worth while, still closer approximations might be secured by using a perfectly conducting cylindrical earth of suitable cross-section.

Thus, the assumption that the return current is confined to the earth's surface does not change the order of magnitude of the effective image distance, but reduces it from about $\frac{3}{4}$ to about $\frac{1}{4}$ of the length of the exposure. For space returns the effective image distance is

$$2AB/e^{3/2}=0.446\,AB.$$

Now take another practical case by assuming that the conductor ab is of negligible length compared with its separation r from the parallel conductor AB and the separation is, in turn, negligible compared with the length AB. The formula for dN/dx given in footnote 6 shows that the required inductance depends only on the ratio $AB/(r_1+r_2)$ and is thus constant upon an ellipse. Equivalent expressions in logarithmic form are

$$N=2\,ab\log\left(2\cos\tfrac{1}{2}\theta_1\sin\tfrac{1}{2}\theta_2\frac{\sqrt{r_1r_2}}{y}\right) \tag{13}$$

$$=2\,ab\log\left(\frac{\cos\frac{1}{2}\theta_1}{\cos\frac{1}{2}\theta_2}\sqrt{\frac{r_1}{r_2}}\right), \tag{14}$$

or approximately

$$N=2\,ab\log\frac{AB}{r},\text{ if } ab \text{ is opposite the midpoint of } AB, \tag{15}$$

$$N=\frac{1}{2}\left[2\,ab\log\frac{AB}{r_2}\right],\text{ if } ab \text{ is at distance } r_2 \text{ beyond } B.[8] \tag{16}$$

Thus, from (13) the effective image distance is never greater than twice the geometrical mean distance from ab to A and B. Its maximum value is approximately AB and occurs when ab is opposite the midpoint of AB. Its minimum value is approximately $\sqrt{r_2AB}$ and occurs when ab is at the distance r_2 from A or B in the prolongation of AB. This makes the inductance one-half of what it is at the symmetrical position. Thus, wherever ab is placed, its mutual inductance with AB lies somewhere between 50 per cent. and 100 per cent. of the mutual inductance, due to an effective image distance AB, ab remaining always parallel to AB and the locus of ab being a rectangle with semicircular ends of radius r_2 and centers A and B.

8. Mutual Impedances of Grounded Circuits Which Depart from the Surface of the Earth

Consider a system of conductors following any paths in space and insulated from the earth except at two grounding points on the sur-

[8] The shortest distance from ab to AB might have been designated by a single letter in place of using y, r and r_2 in formulas (13), (15) and (16).

face of the earth. Any flow of current through this system of conductors may be divided into elementary filaments each of which is made up of segments beginning and ending at the earth's surface and not crossing the earth's surface between these terminals. Ground each segment at both ends. Let $\mathcal{W}$ and $\mathcal{U}$ designate segments having terminals A and B, $\mathcal{W}$ (for example, an open wire circuit) never going below the surface of the earth and $\mathcal{U}$ (for example, an underground cable circuit) never going above the surface of the earth. Add the underground flow $\mathcal{U}$ to Table II at the foot of the left-hand column and, proceeding as before,

$$0 = N_{(\mathcal{U}-\mathcal{X})(z-a)} = N_{\mathcal{U}z} - N_{\mathcal{U}a} + \tfrac{1}{2}\Delta = N_{\mathcal{U}z} + N_{\mathcal{U}o} - \tfrac{1}{2}\Delta,$$

$$0 = N_{(\mathcal{U}-\mathcal{X})(h-a)} = N_{\mathcal{U}h} - N_{\mathcal{U}a} - \tfrac{1}{2}\Delta = N_{\mathcal{U}h} + N_{\mathcal{U}o} - \tfrac{3}{2}\Delta,$$

$$\text{or } N_{\mathcal{U}o} = \tfrac{1}{2}\Delta - N_{\mathcal{U}z} = \tfrac{3}{2}\Delta - N_{\mathcal{U}h},$$

and similarly for the flow $\mathcal{W}$ above ground,

$$N_{\mathcal{W}o} = \tfrac{1}{2}\Delta + N_{\mathcal{W}n} = N_{\mathcal{W}h} - \tfrac{1}{2}\Delta.$$

Hence the three cases which may occur give

$$\left.\begin{aligned} N_{(\mathcal{W}-\mathcal{E})(w-e)} &= N_{\mathcal{W}w} - N_{\mathcal{W}n} - N_{\mathcal{N}w} \\ &= N_{\mathcal{W}w} - N_{\mathcal{W}h} - N_{\mathcal{H}w} + 2\Delta, \end{aligned}\right\} \tag{17}$$

$$\left.\begin{aligned} N_{(\mathcal{U}-\mathcal{E})(u-e)} &= N_{\mathcal{U}u} + N_{\mathcal{U}z} + N_{\mathcal{Z}u} \\ &= N_{\mathcal{U}u} + N_{\mathcal{U}h} + N_{\mathcal{H}u} - 2\Delta, \end{aligned}\right\} \tag{18}$$

$$\left.\begin{aligned} N_{(\mathcal{W}-\mathcal{E})(u-e)} &= N_{\mathcal{W}u} - N_{\mathcal{W}n} + N_{\mathcal{Z}u} \\ &= N_{\mathcal{W}u} - N_{\mathcal{W}h} + N_{\mathcal{H}u}. \end{aligned}\right\} \tag{19}$$

The importance of these equations lies in the fact that the earth return flows are replaced by the simpler nadir, zenith and horizon return flows. *If the conductors comprise only broken straight filaments, making any angles with each other and the earth, the required Neumann integrals, if we use the expressions involving the nadir and zenith returns, are the known expressions between straight filaments.* If the conductors lie in horizontal planes with vertical ground connections, it is convenient to employ the expressions involving the horizon return flows, since the required integral is (4a) derived above. Formulas (33)–(41) of the appendix are the resulting formulas for the three general cases and for a number of important special cases.

In general we may say that the effect of changing the height of one or both conductors by an amount which is small compared with the length of the conductors will be relatively small, since the effective

image distance has been shown above to be of the order of the length of the conductors. To illustrate this fact, in Fig. 8 four dotted curves have been added to the curve of Fig. 5 showing the mutual inductances of the two parallel grounded conductors when they are in the

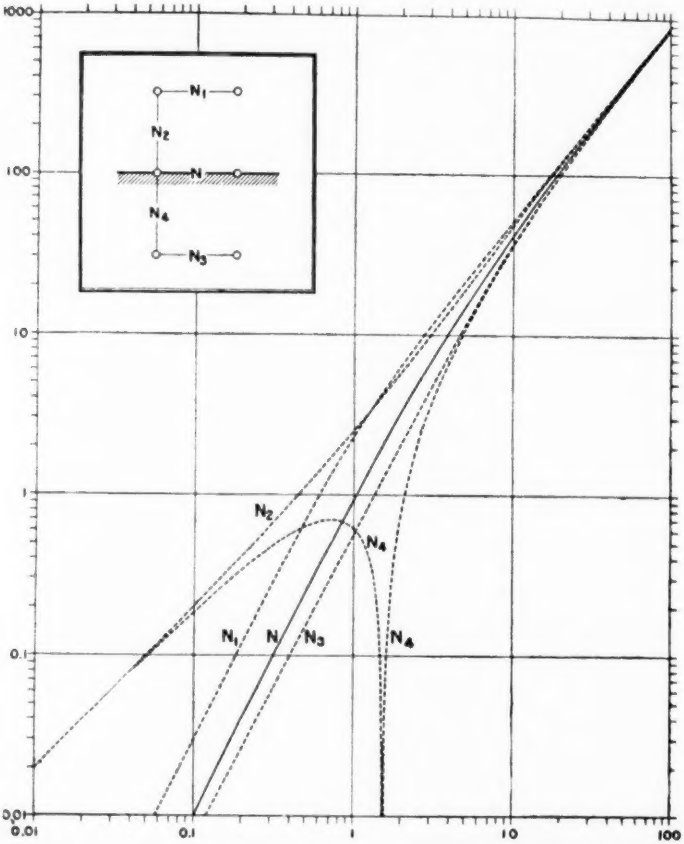


Fig. 8—Mutual Neumann integral between grounded filaments forming opposite sides of a rectangle with unit separation, on the surface of the earth (from Fig. 5) and between the same filaments when one or both of the filaments is raised above or depressed below the surface of the earth as shown by the insert

four positions indicated by N_1, N_2, N_3 and N_4 on the insert of Fig. 8, calculated by formulas (28)–(30) of the appendix, which are special cases of (35), (36), (39) and (40). When the conductors are long, the relative change in the mutual inductance is small. Depressing the

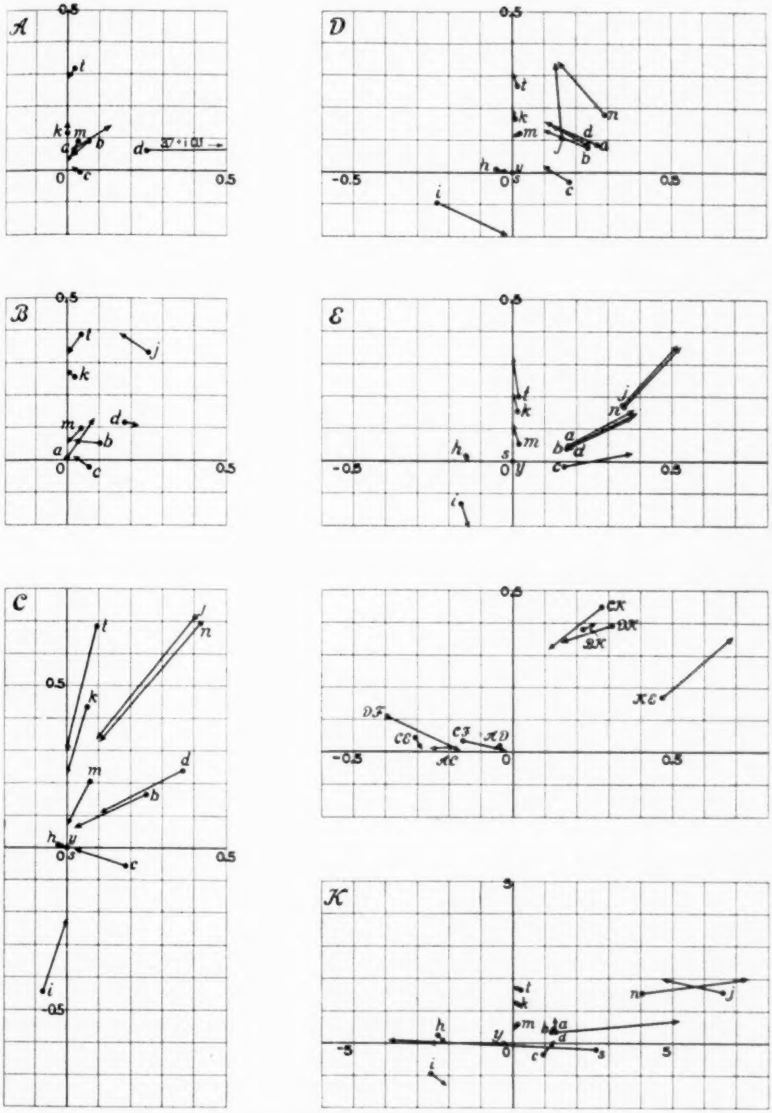


Fig. 9—Comparison of measured and calculated mutual impedances between the indicated circuit pairs. Each arrow extends from the measured impedance located at its tail to the calculated impedance at its head. The location of and positive directions for these circuits are shown by Fig. 10

wires reduces the inductance; the curves show that in the case of N_4 the inductance passes through zero and is reversed in sign when $s = 1.560$.

When the departure of the circuits from the earth's surface may be neglected, all terms, but the first, on the right-hand side of formulas (17)–(19) drop out, and each reduces to the simple, fundamental grounded circuit formula (4).

9. Comparison of Theoretical Results with Measurements at 25 and 60 Cycles

Fig. 9 shows, by means of arrows, the impedances which must be added to each of a large group of measured 25-cycle mutual impedances to obtain the results calculated by means of the preceding formulas, on the assumption that the earth has a uniform resistivity

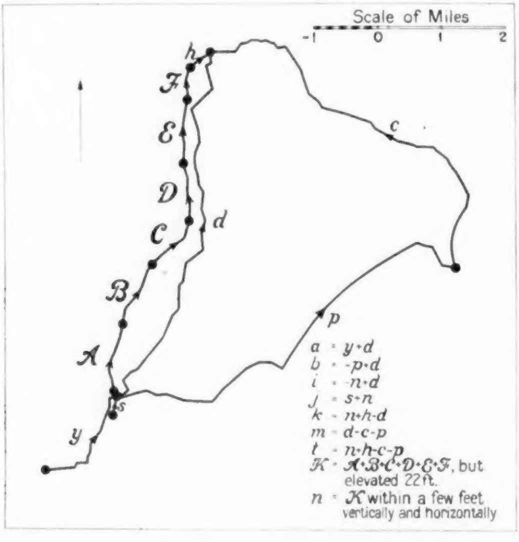


Fig. 10—Location of test circuits, the arrow-heads showing the positive directions in the conductors. When circuits are combined in series, with the removal of intermediate grounds, the new circuit designation is shown by the equations. The test circuits ***k***, ***m*** and ***t*** are large metallic loops from which all grounds have been removed. The horizontal and vertical displacements of the conductor by a few feet, which render the indicated equalities for $\mathcal{K}$ and ***n*** only approximate, were allowed for in determining the calculated results for Fig. 9. The grounds of the capital letter circuits were isolated sections of single track about one mile in length; the midpoint of the section is shown as the effective ground but it may have actually been displaced and have varied with the moisture of the road-bed

of 0.5 megohm per centimeter cube. The measurements were not made for the purpose of this comparison, for which they are not well adapted, but they do give both components of the mutual impedance, which is the absolutely essential requirement. The geometrical irregularity of these circuits is shown by Fig. 10. This was completely allowed for by making detailed computations after substituting an approximately equivalent broken line for each circuit. The variability in the earth's resistivity with location, depth and changing moisture content on different days could not be allowed for. The effect of buried gas and water pipes and of other grounded conductors was also necessarily neglected.

The direct-current theory leaves but one arbitrary constant at our disposal after the frequency and the geometry of the circuits have been fixed. This constant is the earth's resistivity. By trial it was found that 0.5 megohm gave a good average agreement between the calculated resistance components and the entire set of measurements, only a part of which is included in Fig. 9. The individual discrepancies are large but are not so large as to be disconcerting, considering the variations in effective earth resistivity from place to place and from day to day during the progress of the tests.

The calculated reactance component of the mutual impedance based on the direct-current mutual inductance is independent of the earth's resistivity and is uniquely determined by the frequency and geometrical relations. Even a general agreement between the calculated and the measured reactances is significant and Fig. 9 shows not only this, but also a great many good individual agreements. The outstanding discrepancies for circuits $\mathcal{C}$ and $\mathcal{E}$ are systematic, and are apparently to be explained by the effective grounding of these circuits at some other points than the midpoints of the track sections. On the basis of this comparison, it appears that the direct-current theory proves itself adequate to give an approximation to the actual mutual reactances, provided the linear scale and the frequency involved do not greatly exceed those of these tests.

Measurements were also made at 60 cycles. The resistance component remained roughly the same as for 25 cycles; the reactance component doubled as shown by Fig. 11; each component therefore agreeing approximately with the results which would obtain if the direct-current distribution is maintained.

Other comparisons have been made with the same conclusion, but tests should be made, throughout a range of frequencies, at a locality where it is known that the conductivity of the upper layer of the earth's surface is reasonably uniform so that the effect of the lower

layers may be determined. Small scale models, having the proper propagation constant, could advantageously be used in determining the alternating-current impedances for uniform earth resistivity or any assigned distribution of resistivity.

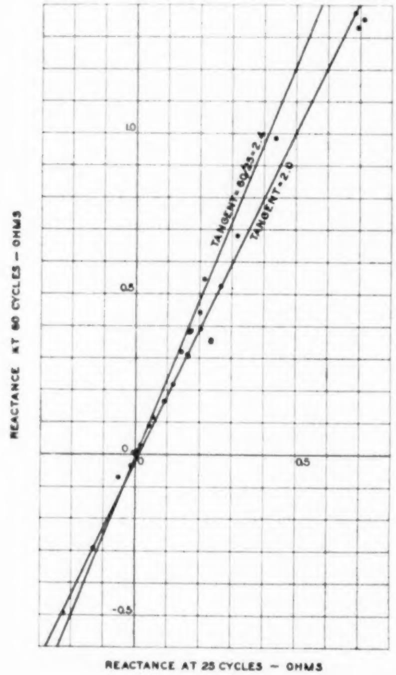


Fig. 11—Comparison of measured mutual reactances at 25 and 60 cycles. A right line with the slope 2.0 fits the point somewhat better than a line with the slope 2.4 which corresponds to the ratio of the frequencies

10. Summary

Formulas for the direct-current mutual resistances and mutual inductances for grounded circuits on the assumption of uniform earth resistivity have been derived and useful diagrams prepared. The applicability of these results as a first approximation to many practical alternating-current cases has been shown.

If, as I hope, this paper is free from ambiguities and errors, it is due to a thorough revision by Mr. R. M. Foster; and I am indebted

to Miss Frances Thorndike for the accuracy attained in the numerous curves of Figs. 1–5 and 8, which should make them of practical value in numerical calculation.

Mathematical Appendix

Additional mathematical results which have been employed in connection with the figures and discussion of this paper are brought together below for convenient reference.

Formulas for Fig. 2

$$2s=\rho+d+1, \text{ if } OA=1,\ Oa=\rho,$$

$$d^2=1+\rho^2-2\rho\cos\theta,$$

$$\sin\theta=d\sin(\theta+\phi),$$

$$\sin\phi=\rho\sin(\theta+\phi),$$

$$\cos\theta=(2s-1)-2s(s-1)/\rho,$$

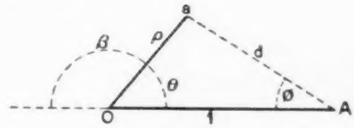


Fig. 12

$$\frac{N\mathscr{R}_r}{AB}=f(\rho,\theta)=\rho f(\rho^{-1},\theta)$$

$$=\cos\theta\left(\log\frac{s}{s-\rho}+\rho\log\frac{s}{s-1}\right) \tag{20}$$

$$=\cos\theta\left\{\log\frac{\tan\frac{1}{2}(\theta+\phi)}{\tan\frac{1}{2}\theta}+\rho\log\frac{1}{\tan\frac{1}{2}\theta\tan\frac{1}{2}\phi}\right\} \tag{21}$$

$$=2\cos\theta\log\frac{2+d}{d}=2\cos\theta\log(1+1/\sin\tfrac{1}{2}\theta), \text{ if } \rho=1,$$

$$=\tfrac{1}{2}\rho\log\frac{2+\rho}{2-\rho}+\tfrac{1}{2}\rho^2\log\frac{2+\rho}{\rho}, \text{ if } d=1,$$

$$=\cos\theta\left[\rho\log\frac{4}{\rho\theta^2}-(1-\rho)\log(1-\rho)\right]$$
$$+\frac{\rho(1+2\rho)}{12(1-\rho)}\theta^2+\ \ldots, \text{ for } \rho<1,$$

$$=\cos\beta[\rho\log\rho-(1+\rho)\log(1+\rho)]-\frac{\rho}{4(1+\rho)}\beta^2$$
$$+\frac{\rho(11+19\rho+11\rho^2)}{96(1+\rho)^3}\beta^4+\ \ldots,$$

$$\theta=\frac{2}{\sqrt{\rho}}e^{-[N+(1-\rho)\log(1-\rho)]/2\rho}+\ \ldots, \tag{22}$$

$$R=\frac{2\rho^2 \log [(1+\rho)/\rho]}{2\log(1+\rho)-\rho/(1+\rho)} = \text{(radius of curvature when } \theta=\pi) \tag{23}$$

$=0,\ 1.564,\ \infty$ at $\rho=0,\ 1,\ \infty$, which checks the sharp curvature of the curves which cross the axis just to the left of the origin.

Formulas for Figs. 4 and 5

$$s=AB/Aa,$$

$$\frac{N_{\mathcal{R}r}}{Aa}=2[s\log(s+\sqrt{s^2+1})+1-\sqrt{s^2+1}] \tag{24}$$

$$\left.\begin{aligned} &=2s\left[\log 2s-1+\frac{1}{s}-\frac{1}{4s^2}+\frac{1}{32s^4}-\frac{1}{96s^6}+\ \ldots\right]\\ &=2s\log\left[\frac{2s}{e}\left(1+\frac{1}{s}+\frac{1}{4s^2}-\frac{1}{12s^3}-\frac{1}{48s^4}+\ \ldots\right)\right]\end{aligned}\right\} \tag{25}$$

$$=s^2\left[1-\frac{s^2}{12}+\frac{s^4}{40}-\frac{5s^6}{448}+\ \ldots\right], \tag{26}$$

$$Q_{\mathcal{R}r}(Aa)\left(\frac{2\pi}{\rho}\right)=2-\frac{2}{\sqrt{s^2+1}}, \tag{27}$$

$$\frac{N_1}{Aa}=\frac{N_{\mathcal{R}r}}{Aa}+2\log(s^2+1), \tag{28}$$

$$\frac{N_2\ (\text{or } N_4)}{Aa}=\frac{N_{\mathcal{R}r}}{Aa}\pm 2[\log \tfrac{1}{2}(1+\sqrt{s^2+1})-\sqrt{s^2+1}+s+1], \tag{29}$$

$$\frac{N_3}{Aa}=\frac{N_{\mathcal{R}r}}{Aa}+2\left[\log\frac{(s^2+1)\ (\sqrt{2}+1)^4}{(1+\sqrt{s^2+2}\ )^4}\right.$$
$$\left.+4(\sqrt{s^2+2}-\sqrt{s^2+1}-\sqrt{2}+1)\right]. \tag{30}$$

Formulas for the Mutual Inductance Between Any Flows in Two Horizontal Planes Grounded by Vertical Filaments

Let the arbitrary flows be $\mathcal{X}'$ and x' between points A', B' and a', b' in the two horizontal planes, grounded by vertical filaments connecting these four terminals with the points A, B, a, b on the surface of the earth. In order to indicate briefly which of these eight points are involved in each term of the result, we imagine a vertical line which cuts the horizontal planes in the points P', p', P, p, where P and p are the same point on the surface of the earth, since the non-

primed points are all in this plane, and we agree that A' or A occurs in a term, according as P' or P is found in the subscript of the symbol Δ or Γ used to designate the term, where

$$\Delta_{P'p'} = -A'a' + A'b' + B'a' - B'b', \tag{31}$$

$$\Gamma_{P'p'} = \log \frac{(A'b' + P'p')(B'a' + P'p')}{(A'a' + P'p')(B'b' + P'p')}. \tag{32}$$

In these expressions every distance between points, such as $A'b'$, $P'p'$, is a positive quantity. The formulas below are perfectly general, but require the assignment of the capital letters $\mathcal{X}'$, A', B', P' to the upper plane when both flows are above the earth and to the lower plane when both flows are below the earth. They are most readily checked by employing formulas (4a) and (24) in formulas (17), (18) and (19). The results show that the mutual inductance is equal to the Neumann integral between $\mathcal{X}'$ and x' augmented by terms which depend only upon the arithmetical distances between the eight points A', B', a', b', A, B, a, b.

$$N(\mathcal{W}-\mathcal{E})(w-e) = N\mathcal{X}'x' + P'p'\Gamma_{P'p'} + 2Pp'\Gamma_{Pp} - \Delta_{P'p'} + \Delta_{Pp}, \quad \text{where } P'p > Pp', \tag{33}$$

$$= N\mathcal{X}'x' + 2Z \log \frac{(Ab)(Ba)}{(Aa)(Bb)}, \text{ if } P' \text{ and } p' \text{ are both at height } Z, \tag{34}$$

$$= N\mathcal{X}'x' + 2Z \log (1+s^2), \text{ if } A'B'b'a' \text{ is a horizontal rectangle and } AB = s(Aa), \tag{35}$$

$$= N\mathcal{X}'x' + 2Z[\log \tfrac{1}{2}(1+\sqrt{1+s^2}) + 1 - \sqrt{1+s^2} + s], \text{ if } A'B'b'a' \text{ is a vertical rectangle with one side on the earth and } AB = s(A'a) = sZ, \tag{36}$$

$$N(\mathcal{U}-\mathcal{E})(u-e) = N\mathcal{X}'x' + P'p'\Gamma_{P'p'} - 2P'p\Gamma_{P'p} - 2Pp'(\Gamma_{Pp'} - \Gamma_{Pp}) - \Delta_{P'p'} + 2\Delta_{P'p} + 2\Delta_{Pp'} - 3\Delta_{Pp}, \quad \text{where } P'p > Pp', \tag{37}$$

$$= N\mathcal{X}'x' - 2Z \log \frac{(Aa)(Bb)(Ab'+Z)^2(Ba'+Z)^2}{(Ab)(Ba)(Aa'+Z)^2(Bb'+Z)^2} + 4(\Delta_{P'p} - \Delta_{Pp}), \text{ if } P' \text{ and } p' \text{ are both at distance } Z \text{ below the earth,} \tag{38}$$

$$= N\mathcal{X}'_{x'} + 2(Aa)t \log \frac{(1+s^2)(t+\sqrt{1+t^2})^4}{(t+\sqrt{1+s^2+t^2})^4}$$

$$+8\,(Aa)\,(\sqrt{1+s^2+t^2}+1-\sqrt{1+s^2}$$

$$-\sqrt{1+t^2}),$$ if $A'B'b'a'$ is a horizontal rectangle at the distance $Z = t(Aa)$ below the earth and $AB = s(Aa)$, (39)

$$= N\mathcal{X}'_{x'} - 2Z[\log \tfrac{1}{2}(1+\sqrt{1+s^2}) - \sqrt{1+s^2}+1+s],$$ if $A'B'b'a'$ is a vertical rectangle with one edge at the surface of the earth and $AB = ab = s(A'a) = sZ$, (40)

$$N_{(\mathcal{W}-\mathcal{E})(u-e)} = N\mathcal{X}'_{x'} + P'p'\Gamma_{P'p'} - 2Pp'\Gamma_{Pp'} - \Delta_{P'p'} + 2\Delta_{Pp'} - \Delta_{Pp}. \quad (41)$$

Thermionic Vacuum Tubes and Their Applications

By ROBERT W. KING

NOTE: The present material was originally prepared for the National Research Council for use in a proposed *Manual* on "Physical Research Methods and Technique." As the appearance of the *Manual* has been postponed, the Committee in charge of its preparation has kindly consented to the separate publication of some of the sections in various technical magazines. In order to meet the requirements of the *Manual*, the form of expression has been made as compact as possible with practically no discussion of theory and no derivation of formulas. Since this style of presentation leaves much to be desired from some points of view, references have been given to the original literature wherever possible. However, many of the vacuum tube circuits presented have not as yet been treated in the literature. In the preparation of the new material the author has been greatly helped by persons whose contact with these subjects is at first hand.

Contents: I. Introduction. II. Two-electrode Tubes. III. Three-electrode Tubes. IV. Thermionic Amplifiers. V. Amplifier Power Supply. VI. Troubles in Amplifier Circuits. VII. Thermionic Modulators. VIII. Thermionic Detectors. IX. Vacuum Tube Oscillators. X. Miscellaneous Applications of Thermionic Vacuum Tubes.

I. Introduction

1. *Thermionic Emission.* By thermionic vacuum tubes we shall understand those whose operation depends in an essential manner upon thermionic emission.

The design of the various types of thermionic tubes at present in use requires no knowledge of the exact mechanism of thermionic emission. It may be said, however, that the work of O. W. Richardson and others leaves little question but that this emission is a physical as distinguished from a chemical process, and occurs from certain substances as the result of the large velocities of thermal agitation acquired by electrons when these substances are raised to a high temperature.

On the basis of certain plausible assumptions, O. W. Richardson derived [2] the expression,

$$I_s = Ne = AT^{\lambda}\epsilon^{-\frac{w_0}{kT}}, \tag{1}$$

for the thermionic emission per cm.^{2} in which I_s is the saturation current formed by drawing all the emitted electrons to a positively charged electrode placed near the emitting surface, e is the electronic charge and ϵ the Naperian base, A is a constant dependent on the emitting substance but independent of the absolute temperature T,

[2] Richardson, The Electron Theory of Matter, 1916 Edition, page 441.

w_o represents the energy lost by each electron as a result of becoming free, λ is a number which does not differ much from unity, and k is the gas constant per molecule. Experiments show that the value of λ for a wide range of substances is about unity, but its exact value is of little practical importance, since the variation of I_s with T is almost entirely controlled by the term in which T enters as an exponent.

The quantity w_o which expresses the *electron affinity* of the emitting substance is usually called the internal work of evaporation. In Equation 1, it is in terms of ergs per electron. Calling v_o the value of w_o when expressed in equivalent volts, $w_o = 1.59 \times 10^{-12}\ v_o$.

The term v_o is of great importance when considering the economy with which a substance acts as a thermionic emitter. Assuming that the emission of an electron occurs when its velocity acquires a value sufficiently high to overcome the potential drop v_o, it is apparent that the smaller v_o, the more copious will be the thermionic emission at any given temperature. For the substances thus far examined, v_o ranges between about two volts and five volts.[3] The two substances whose thermionic properties we shall consider particularly are platinum, coated with a mixture of barium and strontium oxides, and tungsten. For tungsten the value of v_o is approximately 5 volts, and for coated platinum it varies between 1.67 and 2.05 volts. The value of $(v_o)_A - (v_o)_B$ for two substances A and B is equal (except for a small term expressing the Peltier coefficient) to their contact difference of potential.[4]

2. *Thermionic Properties of Filaments.* In designing electron tubes with predetermined characteristics knowledge of the thermionic emissivity of the proposed filament is necessary. This property may be conveniently represented by curves of the type shown in Figs. 1, 2 and 3. The coordinates have been so proportioned [5] that, provided the electronic emission varies with the temperature as indicated by Equation 1 and the thermal radiation from the filament varies as the fourth power of the temperature, then the relation between the thermionic emission and the heating power supplied to the filament is a straight line.

Fig. 1 gives data for tungsten and for coated platinum filaments, Fig. 2 compares thoriated tungsten filament with pure tungsten and Fig. 3 gives data relating to a special coated filament, the core of which consists mainly of platinum alloyed with about 5% of nickel. Since

[3] For a Table of Values of v_o for the materials commonly used see Van der Bijl—Thermionic Vacuum Tube, page 29; also Langmuir, Trans. Am. Electro-chem. Soc. 29,166, 1916.

[4] Richardson, Electron Theory of Matter, 1916, p. 455.

[5] See Van der Bijl, The Thermionic Vacuum Tube, p. 82.

the thermionic emission of tungsten, and thoriated tungsten when freshly activated, do not vary much between different samples, they are given with sufficient accuracy for tube design purposes by single

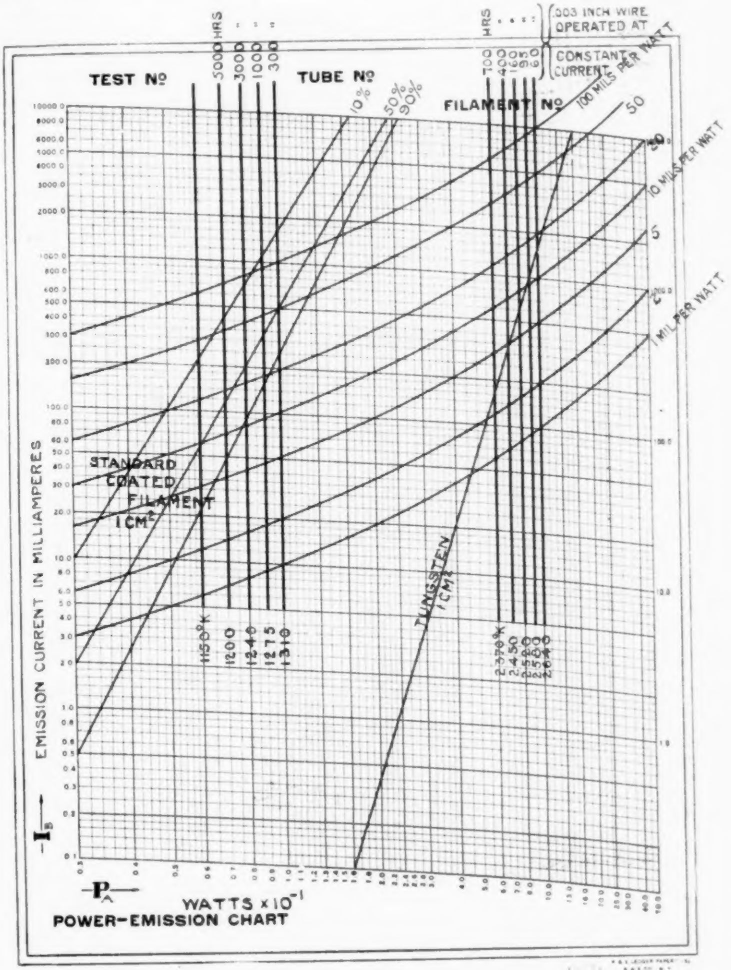


Fig. 1

lines. Coated filaments, however, show a rather wide variation as is indicated. The range of variation shown in Fig. 1 was obtained in the study of several thousand tubes; 10 per cent. showed activity

greater than that represented by the 10% line, 50 per cent. showed activity greater than the 50% line and 90 per cent. showed activity greater than the 90% line. The three lines illustrating the range of

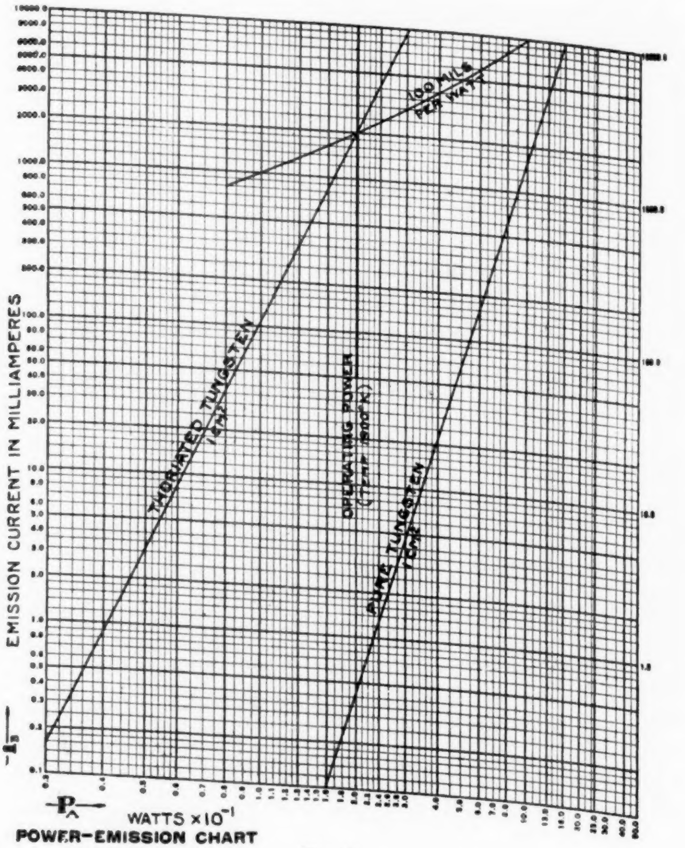


Fig. 2

variation in Fig. 3 do not correspond to these same percentages but have been labeled so as to be readily interpretable.

The efficiencies in milliamperes of emission current per watt used to heat the filament are shown by the curved lines that cross each chart. Operating temperatures and corresponding filament life are given for certain of the ordinates. The tungsten life data in Fig. 1 are for 3-mil wire and a constant heating current. If operating at a con-

stant temperature, the life would be somewhat longer; furthermore, the larger the wire the longer the life for a given temperature. The activity of thoriated filaments tends to diminish with use but may

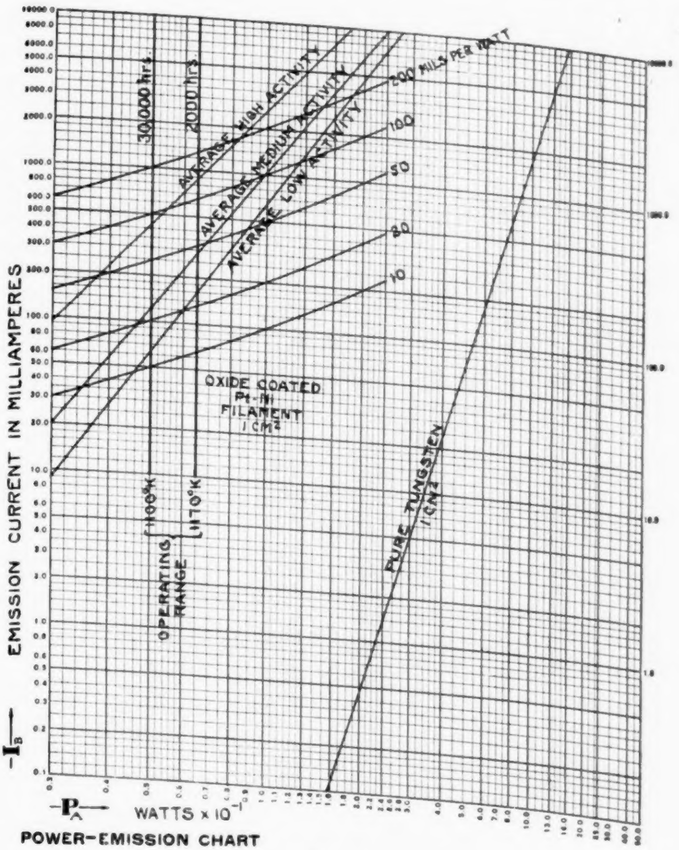


Fig. 3

be re-activated by heating to a temperature higher than the normal operating temperature. The useful life of these filaments is, therefore, somewhat indefinite but taking the possibility of re-activation into account, it is probably well in excess of 2,000 hours.

In using Figs. 1, 2, 3, it should be borne in mind that the emission values given represent saturation currents and that in general the normal operating space current in a tube must, for reasons which will

appear later, be appreciably less than the saturation current. The difference between the saturation current and the maximum operating space current varies with the duty to which the tube is assigned. In the case of high voltage rectifiers, the space current may at certain points in the cycle reach the saturation value, while in a tube which is used as an amplifier it is often desirable, in order to avoid distortion, to have the total emission two to three times as great as the maximum working space current.

3. *Space Current-Voltage Characteristic.* Experiment shows that in a vacuum tube containing an emitting electrode and a conveniently placed anode, the space current I_p, varies with the temperature of the emitter and the anode potential E_p, as in Fig. 4. The three curves shown are for three temperatures such that $T_1 < T_2 < T_3$. It will be observed that between points O and A the three curves coincide;

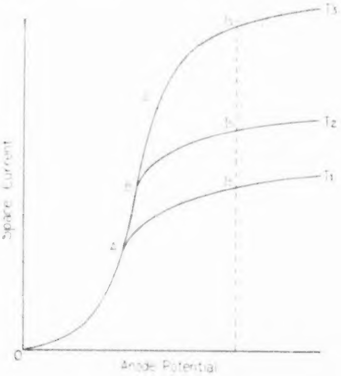


Fig. 4

between O and B the curves for the two higher temperatures coincide. The saturation values of the filament emission at the various temperatures are shown by I_s.

For values of I_p ranging from zero to somewhat less than I_s, the relation between I_p and E_p may be expressed with a fair degree of accuracy by $I_p = \kappa E_p^\eta$ in which the exponent η does not differ greatly from 3/2. This relation, therefore, is frequently known as the 3/2 power law. It has been deduced theoretically by Child [7] and Langmuir [8] and has been studied lately in greater detail by Fry.[9] Fry's analysis takes account of the initial velocities of emission of electrons,

[7] *Phys. Rev.*, Vol. 32, p. 498, 1911.
[8] *Phys. Rev.* (2), Vol. 2, p. 450, 1913.
[9] T. C. Fry, *Phys. Rev.*, Vol. 17, p. 441, 1921.

and, as he shows, the effect of the space charge is to create a region of negative potential immediately around the emitter. Let Fig. 5 represent the value of the potential as one proceeds from the cathode in the direction x, and V' represent the minimum value of the potential. Assuming indefinitely large emission from the cathode, V' (which is a function of E_p), determines the space current corresponding

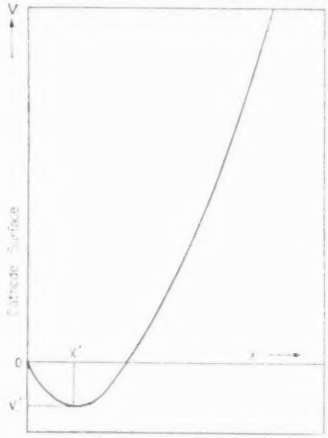


Fig. 5

to any particular value of E_p. The lower the value of V' the fewer the electrons with initial velocities sufficient to carry them past the equipotential surface V' into the region where they are attracted by the anode. Assuming the average initial velocity to be 0.3 volts (roughly a temperature of 2400° K), Fry finds an appreciable deviation from the 3/2 power law for $E_p < 30$ volts, but initial velocities need be considered only in tubes which operate at low E_p.

Another factor which, for low E_p causes an appreciable deviation from the 3/2 power law, is the potential gradient in a filament cathode due to the heating current. Whereas velocity of emission tends to make $\eta < 3/2$, the potential gradient in the filament has the reverse effect. In general, the latter more than overbalances the former and for small anode voltages $\eta > 3/2$. The value of η when the cathode potential gradient is considered, but initial velocities are neglected, has been given by W. Wilson,[10] who finds that when E_p is less than the potential drop across the filament $\eta = 5/2$, while for higher E_p it gradually approaches the limiting value 3/2.

[10] For discussion of the 5/2 power relation, see Van der Bijl, The Thermionic Vacuum Tube, p. 64.

4. *Temperature Saturation and Voltage Saturation.* When a vacuum tube operates at such filament temperature and E_p that an increase in temperature produces no increase in I_p the tube is said to show *temperature saturation.* On the other hand when temperature and E_p are such that an increase in E_p does not increase I_p, the tube shows *voltage saturation.* Referring to Fig. 2 the curve T_1 shows temperature saturation between O and A and approaches voltage saturation beyond the point A. Similarly the curve T_2 shows temperature saturation up to the point B and approaches voltage saturation beyond.

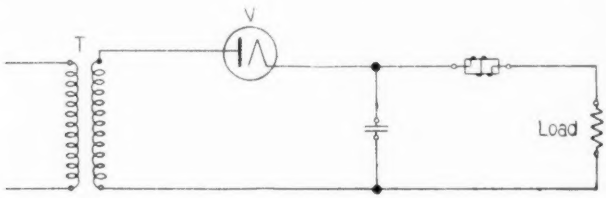


Fig. 6

5. *Effect of Gas.* In thermionic tubes as usually pumped the gas pressure is between 10^{-5} and 10^{-6} mm. At this pressure the gas generally does not manifest its presence in the operation of the tube. However, at higher pressures, and particularly above 10^{-4}mm, it produces certain objectionable disturbances. Thus many gases seriously reduce the filament activity; also for E_p greater than about 20 volts, ionization occurs and the resulting discharge differs in important respects [11] from the pure electron discharge of Fig. 4.

II. Two-Electrode Tubes

The two-electrode tube, which was first due to Edison, found an early practical application when Fleming used it to detect wireless telegraph signals.

However, since the advent of the three-electrode tube of De Forest, the earlier device has been almost entirely superseded as a detector and finds its principal application as a rectifier of a.c. voltages. Its range of applicability in this field is extremely large. With properly designed tubes, Hull [12] has succeeded in rectifying 5 k.w. at a potential of 100,000 volts, and in its transatlantic radio telephone experiments,

[11] See Van der Bijl: The Thermionic Vacuum Tube, p. 86.

[12] A. W. Hull, *General Electric Review*, Vol. 19, p. 177, 1916. Another good source of information is Van der Bijl's "The Thermionic Vacuum Tube."

the American Telephone and Telegraph Company is using a six phase rectifier giving about 200 kw. at about 10,000 volts.[13]

6. *Two-Electrode Tube as Rectifier.* Three typical forms of circuit are shown in Figs. 6, 7, 8, each of which has certain characteristics not possessed by the others. The circuit shown in Fig. 6 rectifies the full transformer voltage, but utilizes only one-half of the current wave; the circuit of Fig. 7 gives a d.c. voltage of only about half the transformer peak voltage but utilizes both halves of the wave, and Fig. 8 illustrates a circuit making use of the full transformer voltage and both halves of the a.c. wave.

A rectifier circuit employing a tuning condenser for the secondary of the high voltage transformer and giving a rectified d.c. voltage as large again as the a.c. voltage of the transformer and providing automatic control of the maximum d.c. voltage supplied by the rectifier, has been described by Webster.[14]

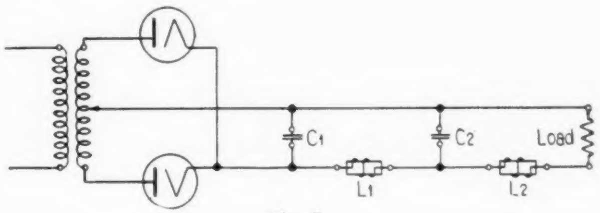


Fig. 7

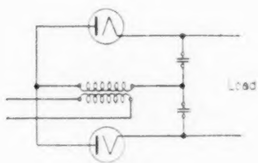


Fig. 8

The circuits shown in Figs. 6 and 7 provide means in the form of condensers C_1, C_2 and inductances L_1, L_2 for smoothing out the rectified voltage. Such an arrangement of conductors is essentially a network whose attenuation for electric currents of a certain range of frequencies is very low and for all other frequencies is high. This type of network is a special form of the *electric wave-filter* which is finding many ap-

[13] Arnold and Espenschied, *Journal of A. I. E. E.*, August, 1923; also *Bell System Technical Journal*, October, 1923.

[14] D. L. Webster, Procd. Nat. Acad., Vol. 6, p. 28, p. 269, 1920. These articles also suggest certain modifications of Hull's method.

plications at the present time and a complete account of its properties is to be found in recent literature.[15]

No definite statement can be made as to the exact range of frequencies over which the rectification of alternating currents can be satisfactorily accomplished by means of thermionic tubes, but it is apparent that this range is large. The degree of smoothness required in the d.c. output is of primary importance in setting the lower limit of the frequency range; on the other hand, the smaller the load resistance, the higher the frequency which may be satisfactorily rectified before the internal capacity of the tubes permits the flow of an objectionable amount of alternating current. Whenever an output with a minimum of ripple is required it is in general desirable to use as high a frequency as is readily available.

III. Three-Electrode Tubes

In 1906 De Forest brought out the three-electrode tube [16] in which a grid is interposed between filament and plate. Since the introduction of this tube, much study has been devoted to its properties and many investigations have been made concerning the best substances to use as thermionic emitters, the best metals for plates and grids, the best varieties of glass for the containing bulbs,[17] and the best methods of exhaustion,[18] so that today problems of design are well understood. At present the three electrode tube finds use as a rectifier, amplifier of small currents and voltages, detector of small a.c. voltages, modulator of alternating currents, and generator of electric oscillations. Tubes have been built which range in size from those about one inch long with a space current of about a milliampere to others which are water-cooled and have an individual output capacity of 100 k.w.[19] Amplifiers with a capacity of 150 k.w. are in operation (see footnote 13).

7. *Action of the Grid.* The general theory of the grid action is simple. As pointed out by Fry [19] the space charge creates a region of

[15] G. A. Campbell, *Bell System Technical Journal*, Nov., 1922; U. S. Patents 1,237,113 and 1,237,114, May 22, 1917; O. J. Zobel, *Bell System Technical Journal*, Jan., 1923; Carson and Zobel, *Bell System Technical Journal*, July, 1923; G. W. Pierce, Electric Oscillations and Electric Waves, p. 186, 1920; Karl W. Wagner, *Arch. f. Electr.*, Vol. 3, p. 315, 1915; Vol. 8, p. 61, 1919.

[16] Variously called the audion, vacuum tube, triode, pliotron, etc.

[17] Measurements of Gases Evolved from Glasses of Known Chemical Composition—Harris & Schumacher, *Jour. Ind. & Eng. Chem.*, Feb., 1923; also *Bell System Technical Journal*, Jan., 1923.

[18] For methods of exhaustion, see Dushman, *Gen. Elect. Review*, Vol. 23, p. 493, 1920, et seq.

[19] See W. Wilson, *Bell System Technical Journal*, July, 1922.

[19] T. C. Fry, l. c.

negative potential immediately around the cathode which persists for all values of E_p less than that required to produce voltage saturation. It is the minimum potential V' (see Fig. 5) that limits I_p, and any increase in V' will result in an increase in I_p. Because the grid is close to the filament, small changes in the grid potential, E_g, are as effective in changing V' and therefore I_p, as large changes in E_p. This leads to the so-called amplification constant μ of the tube which may be taken as

$$\mu = \frac{\pm e_p}{\mp e_g},$$

in which $\pm e_p$ and $\mp e_g$ are changes in E_p and E_g, the changes being opposite in sign as indicated, and such that they leave I_p unchanged.

It has also been shown[20] that if ΔE_p and ΔE_g are increments of E_p and E_g which cause equal increments in the electric field at the surface of the cathode (considered simply as an equipotential surface and not as a source of electrons) the amplification constant, μ, of the tube will be the ratio $\Delta E_p / \Delta E_g$.

8. *Characteristic Equation.* The electrical characteristics of the three-electrode vacuum tube may be represented [21] by the equation

$$I_p = \kappa\left(\frac{E_p}{\mu} + E_g + \epsilon\right)^{\eta}. \qquad (2)$$

The constant κ is related in a simple way to the internal resistance of the tube. A consideration of ϵ which expresses the contact difference of potential between grid and filament is usually essential only in tubes which operate at a low E_p and particularly in detectors and amplifiers. In tubes with coated filament, ϵ may not only vary within a range of two or three volts between different tubes, but may also change during the life of any one tube. The exponent η varies in a given tube with applied voltage, being usually equal to about 2 when the effective voltage $\left(\frac{E_p}{\mu} + E_g + \epsilon\right)$ is comparable with the potential drop in the filament (see Fig. 9), and tending to approach the theoretical value 3/2 with increasing effective voltage. It has been found possible to deduce relations between the constants μ and κ of Equation 2 and the structure and dimensions of any tube which are in very fair agreement with experimental values.[22]

Typical curves corresponding to the characteristic Equation 2 are shown in Figs. 10 and 11. These curves are referred to as *static* char-

[20] R. W. King, *Phys. Rev.*, Vol. 15, p. 256, 1920.

[21] Van der Bijl, *Phys. Rev.*, 12, 180, 1918.

[22] King, l. c.

acteristics one parameter being fixed in each case. For the *dynamic* characteristic see section 12. Equation 2, of course, fits only the portions of the curves characterized by temperature saturation.

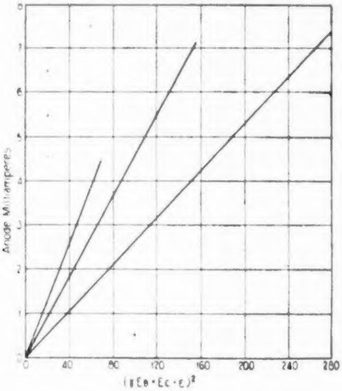


Fig. 9

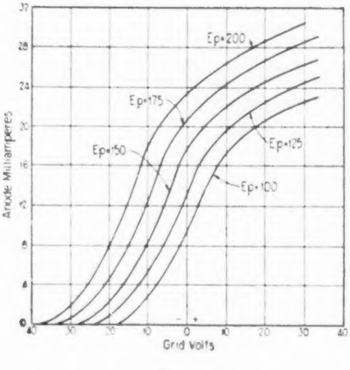


Fig. 10

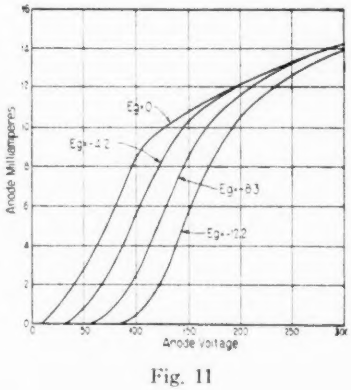


Fig. 11

The abruptness with which the curves of Fig. 10 meet the potential axis is important in certain uses to which tubes are put. The value of E_g which reduces I_p to zero is called the *cutoff voltage*. To have a sharp cutoff a tube should have a fairly large μ and its grid should be sufficiently large with respect to the filament to effectively screen all parts of the filament from the plate.

9. *Grid Current.* For certain purposes, a consideration of the grid current I_g, is necessary. Fig. 12 represents a characteristic relation between I_g and E_g for various E_p. Note that E_g in excess of about 10 volts results in *secondary emission* of electrons from the grid. These secondary electrons flow to the plate and, as shown, their number may actually exceed the total number of primary electrons striking the grid. The character of the grid surface plays an important part

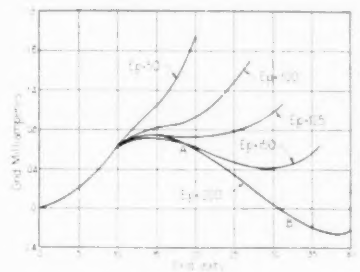


Fig. 12

in determining the amount of secondary emission. The secondary emission from the grid of a tube containing a pure tungsten filament is, in general, less than that from the grid of a tube with an oxide-coated filament. At high temperature a coated filament appears to evaporate a minute amount of its coating, some of which is deposited upon the grid presumably augmenting the secondary activity.[23]

10. *Vacuum Tube Constants.* The two most important constants of the three electrode tube are μ and its internal resistance r_p. The determination of μ and r_p from the characteristic curves (Figs. 10, 11) is obvious. For general design purposes these curves give the best insight into the behavior of a tube and furnish the most instructive means of determining μ and r_p.

11. *Dynamic Methods of Measuring Vacuum Tube Constants.* However, in cases where many tubes, all practically alike, have to be tested, certain "dynamic" methods are timesavers. Several such methods have been devised, but all are modifications of a scheme first published by Miller.[24]

[23] A. W. Hull has designed two types of tube known as the dynatron and pliodynatron which utilize the negative resistance characteristic (AB of Fig. 11) resulting from secondary emission. Procd. Inst. Radio Engrs., Vol. 6, p. 5, 1918.

[24] J. M. Miller, Proc. I. R. E., Vol. 6, p. 141, 1918. For variations of Miller's dynamic method the reader is referred to Van der Bijl, l. c., p. 198, Method of G. H. Stevenson.

Miller's method is illustrated in Fig. 13. To determine μ the key K_1 is open and K_2 is closed. The resistance r_1 is adjusted until the sound heard in the telephones T is a minimum, under which circumstances it is clear that $\mu = \frac{r_1}{r_2}$. To determine r_p, some definite relation between r_1 and r_2 is established. Then, with key K_1 closed, the resistance r_o is adjusted until the telephone response is a minimum. With this adjustment it may be shown that

$$r_p = r_o\left(\mu\frac{r_2}{r_1} - 1\right). \tag{3}$$

This measurement of r_p may be simplified as follows: suppose we adjust r_1 for a minimum tone in T when K_1 is open. Then $\mu = \frac{r_1}{r_2}$, and it is seen from Equation 3 that with this relation between r_1 and r_2 it would not be possible to obtain a balance with K_1 closed; but if

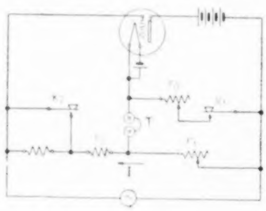


Fig. 13

r_2 be doubled, which can be done by opening K_2 thus adding a resistance equal to r_2, and r_o be now adjusted with K_1 closed to give a minimum tone in T, then $r_p = r_o$.

12. *Dynamic Characteristics of Vacuum Tubes.* In a circuit such as that shown in Fig. 13, the space current causes a fall of potential along any resistance r, and the difference in potential between filament and plate is therefore less than the potential difference across the battery by the amount $I_p r$. If I_p is increased in any way, as for instance, by an increase in E_g, the drop $I_p r$ increases and with a fixed battery e.m.f. the potential difference between the filament and plate diminishes somewhat. It follows, therefore, that a given change in E_g will cause a smaller change in space current when the plate circuit includes an external resistance r than when it does not.

This important fact supplies a simple means of straightening the characteristic of a vacuum tube to such an extent that it may become practically a distortionless amplifier.

To a first approximation,[26] the alternating component J of the space current, when a voltage $e=e_o \cos pt$ is applied to the grid is given by

$$J=\frac{\mu}{r+r_p}e_o \cos pt-\frac{r_p r_p' e_o^2}{2!(r+r_p)^3}(1+\cos 2pt). \tag{4}$$

In this equation r_p is the internal resistance of the tube and r_p' is its derivative with respect to the effective voltage $\left(\frac{E_p}{\mu}+E_g\right)$. It will be noted that the second term on the right side of Equation 4, which gives the first harmonic, diminishes rapidly with r as was to be expected from the preceding paragraph. The more or less straightened

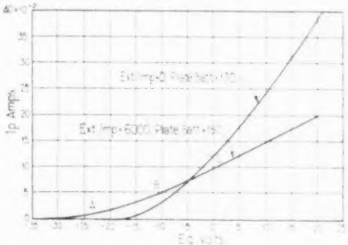


Fig. 14—Average I_p, E_g Characteristics for Five Tubes. Average $\mu=5.92$; average internal impedance $=6,000$ ohms for $E_b=130$ volts and $E_g=-9$ volts. Plate battery connected to +end of filament and grid battery to −end. Western Electric type 101-B tubes.

characteristic resulting from the effect of r is known as the *dynamic* characteristic; see Fig. 14, which curves fit a tube whose $r=6{,}000$ ohms. As will be pointed out in the section on amplifier circuits, the dynamic characteristic is a useful guide in selecting tubes as amplifiers.

Equation 4 also expresses the important fact that the application of the voltage e_o to the grid is equivalent, so far as current in the plate circuit is concerned, to the application of the voltage μe_o in the plate circuit.

13. *Internal Capacities and Effect on Input Impedances.* In certain uses to which the vacuum tube may be put, a knowledge of the influence which the internal electrostatic capacities have on the input impedance is important. The equivalent circuit of the tube[27] is

[26] J. R. Carson, Proc. Inst. Radio Engrs., Vol. 7, page, 187, 1919. In case the output circuit of a tube contains reactance as well as resistance, Eq. 4 becomes much more complicated as Carson shows.

[27] H. W. Nichols, *Phys. Rev.*, Vol. 13, p. 405, 1919; J. M. Miller, Bureau of Standards, Bulletin No. 351.

shown in Fig. 15 in which C_1 is the capacity between filament and grid, C_2 capacity between filament and plate, C_3 capacity between grid and plate, and r_1, r_3, are leakage resistances. As the action of the tube is such as to produce an equivalent voltage μe_o between filament and

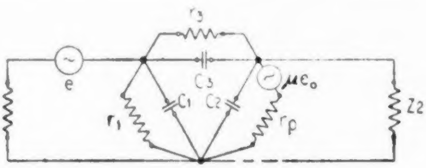


Fig. 15

plate, a generator of voltage μe_o is shown in series with the internal resistance r_p of the tube. Calling Y_g the input admittance of the tube, that is,

$$\frac{1}{Y_g}=Z_g=\frac{e_g}{i_1+i_3},$$

in which e_g is the alternating voltage between filament and grid, the solution of the above circuit gives for Y_g the value,

$$Y_g=\frac{1}{r_1}+jC_1p+\frac{\left(\frac{1}{r_3}+jC_3p\right)[jC_2pr_pZ_2+r_p+Z_2(\mu+1)]}{(jC_2pr_pZ_2+r_p+Z_2)+\left(\frac{1}{r_3}+jC_3p\right)r_pZ_2}. \tag{5}$$

Case 1, *Low Frequencies.* For low frequencies the admittance of the condenser C_2 is negligible compared with that of r_p. Dropping the terms containing C_2 gives the equation,

$$Y_g=\frac{1}{r_1}+jC_1p+\left(\frac{1}{r_3}+jC_3p\right)\frac{r_p+Z_2(\mu+1)}{(r_p+Z_2)+r_pZ_2\left(\frac{1}{r_3}+jC_3p\right)},$$

which yields important interpretations. In case the load impedance Z_2 is a pure resistance r_2, the admittance of the filament-grid branch of the tube may be much greater than the admittance which would result from R_1 and C_1 alone. This is due to the influence which the alternating component of the plate voltage exerts upon the input circuit through the condenser C_2. Figs. 16 and 17 show respectively the effective capacity and effective conductance between filament and grid as a function of the external resistance. For the particular tube studied (W. E. Co. 102-A) Fig. 17 shows that, if $r_2=40{,}000$ ohms, the effective capacity between filament and grid is not the capacity C_1

(about 10×10^{-12} farads) but is approximately 120×10^{-12} farads. Fig. 16 shows that the effective conductance is also greatly increased above that due to r_1. This increase in conductance means an increased absorption of input energy by the tube which, of course, is

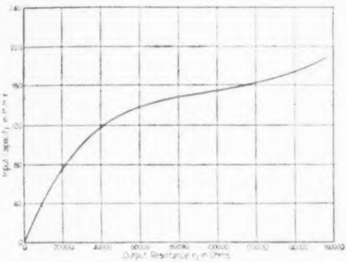


Fig. 16

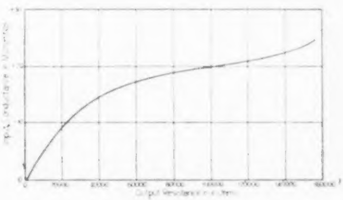


Fig. 17

not dissipated in the grid circuit but passes through the path supplied by the condenser C_3 to be wasted in the plate circuit.

In case Z_2 is a pure inductance L_2, the effective input conductance of the tube is negative and not positive as in the preceding case.

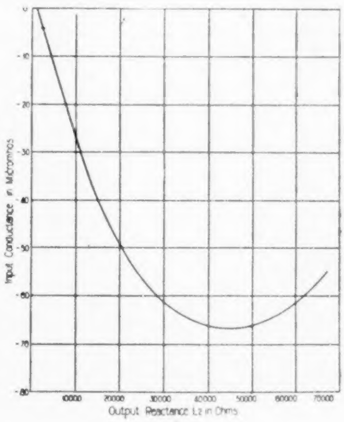


Fig. 18

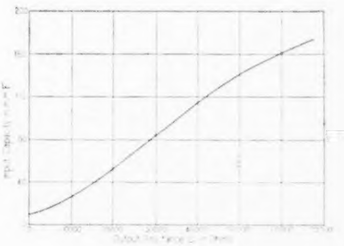


Fig. 19

Fig. 18 shows the variation of this negative conductance with L_2, and Fig. 19 the variation of effective input capacity with L_2. A negative input conductance means that the grid circuit draws power from the plate circuit; if the negative conductance is large enough, a tube in

such a circuit will oscillate steadily or "sing" with no coupling but that provided by its internal capacities. This phenomenon is frequently encountered in vacuum tube amplifiers and at times proves quite troublesome.

Tubes can readily be constructed in which r_1 and r_3 are so large as to exert no influence on the behavior of the tube and may be ignored in the above equations. However, even in such tubes there is an effective input conductance, either positive or negative, depending upon the character [28] of Z_2.

Case 2, High Frequencies. For very high frequencies terms of the first order in p are negligible compared to terms of the second order, and Eq. 5, becomes,

$$Y_g = \frac{p(C_1C_2 + C_1C_3 + C_2C_3)}{C_2 + C_3},$$

indicating that as the frequency is raised the effective input impedance approaches that due to the condensers alone. Under these circumstances the grid absorbs very little power, but the amplification is lowered because the input is to an extent short-circuited by the electrode capacities. Fig. 20 shows the variation in voltage am-

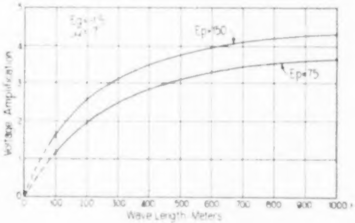


Fig. 20

plification against wave lengths in meters for high frequencies. The two curves are for different E_p, the higher E_p giving a larger amplification because r_p of the tube is lower. It is seen that the amplification at 1,000 meters is about three times as large as the amplification at 100 meters and the amplification at both values of E_p tends to approach zero as the frequency becomes infinite.

Nichols suggests [29] that the reduction in amplification for a given frequency can be avoided by shunting the grid-plate capacity C_3 with

[28] For cases in which Z_2 is neither pure resistance nor reactance, see Van der Bijl, l. c., p. 210.

[29] H. W. Nichols, *Phys. Rev.*, Vol. 13, p. 411, 1919.

an inductance of such a value as to make the impedance between grid and plate infinite at this frequency.

IV. Thermionic Amplifiers

Equation 4, given in Sec. 12 is of fundamental importance in the design of vacuum tube amplifier circuits. Neglecting the second term on the right hand side which, as previously pointed out, expresses distortional effects and should therefore be very small in amplifier circuits, the equation can be written

$$J = \frac{\mu\, e_o}{Z + r_p}, \tag{6}$$

in which the impedance $Z = r + j\,x$ is substituted for r. From Equation 6 both the voltage and power amplification of a tube for any particular circuit can readily be calculated.

14. *Voltage Amplification.* Assuming as above that the tube works into an output impedance Z, it follows that the voltage amplification (i.e., the ratio of the output to the input voltage) is

$$\frac{\mu\, Z}{Z + r_p}.$$

This expression shows that the voltage amplification increases as Z increases. Considering separately the two cases in which Z is a pure resistance and pure reactance, typical values of the voltage amplification are plotted in Fig. 21. Curve a corresponds to reactance

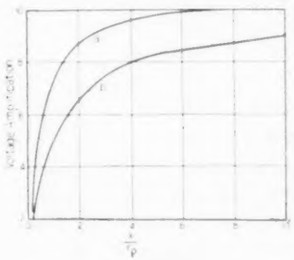


Fig. 21 .

in the output and b to resistance. These curves show that the voltage amplification rises much more rapidly when reactance is used, reaching 90% of its maximum value when $\frac{x}{r_p} = 2$.

If the resistance component of Z is made as small as possible, E_p becomes practically equal to the potential across the plate battery,

making possible a given voltage amplification with smaller plate battery than could be obtained if the output circuit contained an appreciable resistance.

15. *Power Amplification.* From Equation 6, which neglects harmonic terms, it is readily seen that the power output is

$$\frac{\mu^2 e_o^2 r}{(r+r_p)^2+x^2},$$

where $r+jx$ has been substituted for Z. This output is a maximum when $r^2=r_p^2+x^2$. The case in which $x=0$ is particularly important; evidently for maximum power output, a tube should work into a resistance equal to its internal resistance.

As pointed out in Sec. 13 the input impedance of a tube is not always readily determinable; however, calling the input resistance [30] r_g, the power amplification produced by a tube is given by the expression,

$$\frac{\mu^2 r r_g}{(r+r_p)^2+x^2}. \tag{7}$$

This expression has been obtained on the assumption that the grid draws no electron current from the space charge, which in turn requires that the grid remain at a negative potential at all times. Since the power amplification falls rapidly as the grid becomes positive, it is customary in most amplifier circuits to supply means of maintaining the grid at a negative potential.

Expressions 6 and 7 are of fundamental importance in the design of amplifier circuits.

16. *Selection of Tubes.* When selecting tubes for an amplifier, curves such as those shown in Fig. 14 are very useful. By their means it is readily possible to select the tube, the plate potential and the average grid potential which will give satisfactory results for any pre-assigned value of the input voltage. In order to obtain amplification as free from distortion as possible, it is necessary that the grid potential in its excursions neither become positive nor strike the lower end of the characteristic. To a sufficient approximation it is evident that when the variable grid or input voltage e_o is given, we should choose E_g and E_p such that

$$e_o \ngtr -E_g \ngtr \frac{E_p}{2\mu}.$$

[30] Where many tubes of the same design are to be interchanged in a given circuit, and where the conditions of manufacture are such that the insulation resistance between filament and grid is not always of the best, it may be found desirable to shunt the input with a fixed resistance, e.g., ½ megohm.

Referring to Fig. 14, for an input potential of 10 volts (peak value) and $r_p = 6{,}000$ ohms, the point B (but neither A nor C) is evidently a satisfactory mean position about which to operate.

Since both voltage and power amplification increase as E_p is increased (because r_p is decreased) it is frequently desirable to have the value of E_p considerably larger than the lower limit just indicated.

17. *Amplifier Circuits.* The fundamentals of thermionic amplifier circuits may be gathered from Fig. 22. The variable input voltage

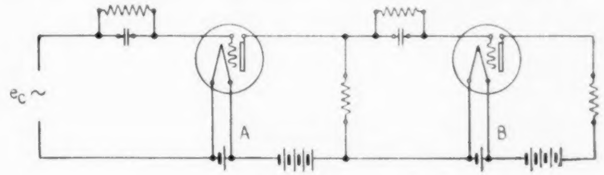


Fig. 22

e_o between grid and filament of the first tube A modulates the plate current of this tube. The circuit is evidently such that the variable I_p of tube A varies the grid potential of B and, due to the properties of the three electrode tube, not only is the power applied to the grid of B greater than that applied to A, but the potential variation may be many times as large as e_o. Hence the variations in I_p of B will be larger than those of I_p in A, thus yielding an amplifying action. Tubes A and B and their associated circuits are known as the first and second stage respectively.

Amplifier circuits may for convenience be divided into six general classes:

1. Resistance coupled circuits (Fig. 23).
2. Resistance-condenser coupled circuits (Fig. 24).
3. Retard-condenser coupled circuits (Fig. 25).
4. Transformer coupled circuits (Fig. 26).
5. Feed-back circuits (Fig. 31).
6. Push-pull circuits (Fig. 32).

The Sections immediately following will point out the advantages of each type of circuit and general design considerations. Equations 6 and 7 show that the amplification of which a single tube (or "stage") is capable is definitely limited. For greater amplification than a single stage can produce, it is necessary to arrange two or more stages in cascade. Multistage amplifiers frequently consist of combinations of certain of the above types of circuits as will be pointed out in the following paragraphs.

18. *Resistance Coupled Amplifier.* This simplest of all amplifier circuits (Fig. 23) is particularly useful where a wide range of frequencies is to be amplified without selective amplification of any particular frequencies. For this reason, it is often used as an amplifier in con-

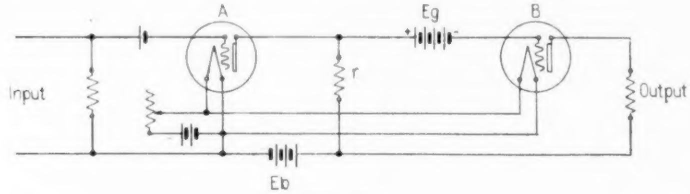


Fig. 23

nection with an oscillograph. It is also one of the few types of circuit which can be used for direct current amplification. However, as will be pointed out later, a special type of push-pull circuit makes a more satisfactory d.c. amplifier for many purposes.

One or more stages of the resistance coupled circuit may be substituted for transformers in voltage amplification. Voltage amplifier tubes having an amplification constant $\mu = 30$ are common and it follows from Equation 6 that such a tube can readily produce a voltage amplification of from 20 to 25. It can, therefore, take the place of an input transformer in one of the other types of amplifier circuit. Unless special considerations require another adjustment, it is customary to arrange all but the last stage of an amplifier for voltage amplification, the last stage being designed for power amplification. (See Secs. 14 and 15.)

Since in resistance coupled amplifiers there is a d.c. path between the plate of one tube and the grid of the following tube, a negative grid battery large enough to counterbalance the plate battery must be used in every stage in order to supply the necessary negative grid potential. As shown in Fig. 23, a common plate battery can be used for two or more stages. A more complete discussion of battery requirements is given under Power Supply.

19. *Resistance-Condenser Coupled Amplifier.* This type of circuit (Fig. 24) is similar to the preceding in all respects except that condensers are inserted between the plates and grids of adjacent stages. This makes the employment of large negative grid batteries unnecessary although it is still important that steady negative potentials be applied to the grid of each tube sufficiently large to prevent their being carried positive by the variations. For example, in Fig. 24, r_g' may be two meghoms and the grid battery emf 2 to 3 volts. This

type of circuit is in general the easiest to handle. Due to the insertion of condensers it will, of course, not serve as a direct current amplifier,

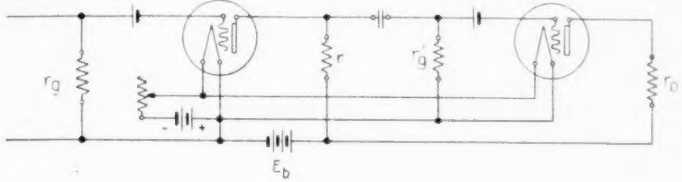


Fig. 24

but with sufficiently large condensers it can handle low frequencies with little or no distortion.

20. *Retard-Condenser Coupled Circuit.* The substitution of retard coils for resistances (see Fig. 25) in the circuit last described affects

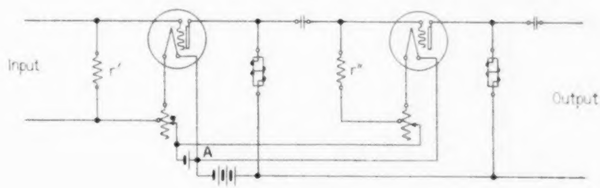


Fig. 25

the behavior of the circuit in several ways. An advantage in the change lies in the fact that a given plate potential can be secured by a smaller plate battery (Sec. 14). Since the tubes in all but the last stage of an amplifier generally act as voltage amplifiers, it is desirable that the inductance of the retard coils be made large. By employing retard coils of the proper inductance and resistance and shunting them with condensers, such an amplifier may be tuned for any particular frequency.

It follows that the width of the frequency band which can be amplified without distortion is likely to be less than for the resistance coupled amplifier. Since the impedance of the retard coils increases with increase of frequency the higher frequencies will, in general, be amplified more than the low. However, it is impossible to make retard coils without a certain amount of distributed capacity, the shunting effect of which tends to limit the amplification of the higher frequencies. By the proper design of coils it is possible to construct a retard-coupled amplifier which will give practically uniform amplification, e.g. throughout the speech range of 200 to 3,000 cycles. It

is customary to make the retard and choke coils of the toroid or closed core type.

21. *Transformer Coupled Amplifiers.* From a theoretical point of view the transformer coupled amplifier (Fig. 26) should be the ideal type. By the proper choice of transformers it should be possible to match stages with respect to one another in such a way as to obtain the greatest efficiency from tubes and batteries. The chief advantage of transformer coupling lies in the fact that the input voltage to the second stage may be made greater than the voltage

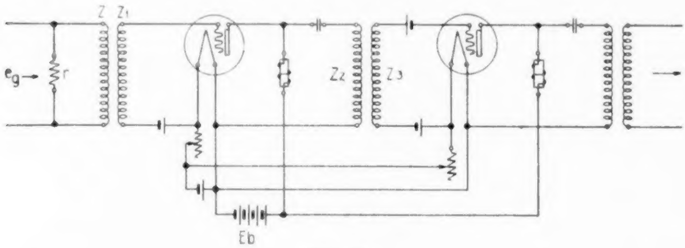


Fig. 26

output of the first, and so on, at the same time that each tube operates into a properly matched impedance to give maximum power output. When uniform amplification over a relatively wide band of frequencies is not required, the interstage transformer may be designed to step up the voltage as many as 30 to 40 times. Other advantages are the economical use of plate batteries (Sec. 14) and the elimination of grid condenser and grid leak or high voltage grid battery.

However, the difficulties attendant upon the design and making of transformers are such that to realize the apparent advantages of this type of circuit will require very careful planning. This will be illus-

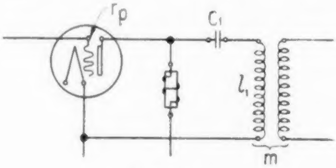


Fig. 26a

trated by the following example of a transformer to handle the frequencies of speech. Given an interstage transformer as in Fig. 26a, we will assume that the transformer works out of a tube impedance r_p and into a grid circuit impedance which has infinite resistance.

Then, imaging for the moment that the condenser C_1 has been removed, the output voltage e_2 of the transformer is

$$e_2 = \frac{emp}{r_p + jl_1 p}, \tag{8}$$

in which e is the input voltage, l_1 is the inductance of the primary winding, m is the mutual between the windings, and p is 2π times the frequency. This neglects resistance of the winding and also capacity effects. Inspection of Equation 8 shows that e_2 varies with the frequency or with p in the manner shown in curve A, Fig. 26b, from which it is seen that the transformer tends to suppress the lower frequencies.

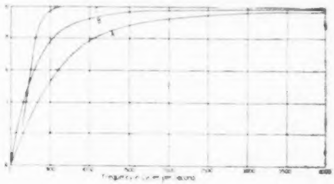


Fig. 26b—Curve A corresponds to $l=1$ henry. Curve C to $l=2$ henries, $C_1=.141$ f.

Curve A shows the performance of a transformer as calculated from Equation 8 assuming $r_p = 5{,}000$ ohms and $l_1 = 1$ henry. Such a transformer would be quite unsuited for a speech frequency amplifier as it introduces very serious distortion below 1,000 cycles. Curve B is calculated on the assumption that $l_2 = 2$ henries and shows marked improvement over A for the lower frequencies.

In input transformer design it is ordinarily necessary to limit the inductance of the two windings not only because of the limited winding space but also because of the need of keeping down the capacity between windings and the capacity within each winding. Curve C shows the performance of the same transformer as Curve B when the capacity C_1 (Fig. 26a) is put in the primary circuit, C_1 having a value of .141 m.f. and being so chosen as to tune l_1 to 300 cycles. With the capacity present Equation 8 becomes

$$e_2 = \frac{e\,m\,p}{r_p + j\left(l_1 p - \dfrac{1}{C_1 p}\right)}. \tag{8A}$$

Use of the capacity improves the transformer characteristic for all frequencies above about 200 cycles and the combination therefore gives better results in a speech amplifier than the transformer alone.

The effect of distributed capacity in the windings (present especially in the secondary because of its greater number of turns) is, more or less, to shunt the high frequencies. This may be counteracted either by the inductance in the primary winding or, if this is not sufficient, by insertion of the proper inductance in series with the primary. It may be said, in a general way, that the lower the ratio of a transformer the better suited its frequency characteristic will be to a wide band of frequencies such as occurs in speech, and transformers with a ratio of 1 to 4 are made which require no correcting provided they are properly chosen with respect to the impedance characteristics of associated tubes.

The selective amplification of an amplifier for particular frequencies may be increased by tuning one or more of the secondaries of the interstage transformers with condensers. (See Equation 8A.)

Due to the fact that there is an appreciable distributed capacity between the primary and the secondary windings, an interstage transformer supplies capacity coupling as well as inductive coupling between adjacent stages, the phase of the capacity coupling being independent of the direction of winding while the inductive coupling is not. Therefore, the transformer may be so placed in the circuit that these two effects either aid or oppose one another. In order to secure the greatest amplification they should aid.

The transformer used in speech frequency work is, in general, made with an iron core; therefore, care should be taken to prevent the d.c. component of I_p magnetizing the core and reducing its efficiency. One method of accomplishing this is shown in Fig. 26 in which the d.c. component of I_p is bypassed by a choke coil. In circuits in which two or more interstage transformers are used, attention should be paid to the danger of magnetic feed-back. This can be largely eliminated by using transformers with closed magnetic circuits. Both the toroid type core and the shell type core (commonly employed in power transformers) have been found satisfactory, and especially the latter.

22. *Amplification of Higher Frequencies.* The transformer coupled amplifier is the type perhaps best suited to use at frequencies higher than those of speech. As a special case the amplifier circuit of Fig. 27 will be considered first. This circuit contains a tuned output and should be used only in case a single frequency or very narrow band of frequencies is to be amplified but in this case will be found very satisfactory. The inductance L may consist of two parallel windings, insulated from each other to avoid any conductive connection between the plate battery and the output. Such an arrangement would be desirable if the output went to a detecting tube or another

amplifying tube. By using a variable condenser C, the frequency of maximum amplification can be readily shifted but for any one setting the amplification will be as shown by curve A of Fig. 28. For maxi-

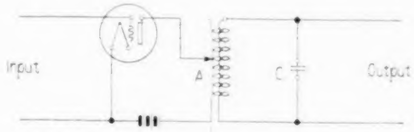


Fig. 27

mum output, the position of the tap A should be set so that the impedance of the tuned circuit LC as seen from the tube is equal to the output impedance of the tube. Tuned circuit amplifiers are especially adapted for amplification at very high frequencies (above 2,000,000 cycles), where the effect of the capacities between the elements of the tubes makes other types of amplifiers very inefficient. The amplification curve A will be broadened when more and more turns are added to the inductance L and the capacity of the tuning condenser C is diminished due to the effect of the distributed capacity

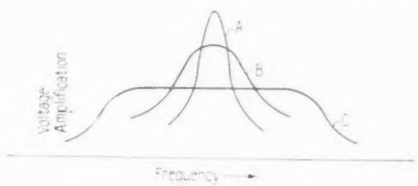


Fig. 28

of the coil. Curve B gives the amplification for a circuit where the condenser C is omitted in which case it becomes a retard coupled amplifier. Maximum amplification occurs at the natural frequency of the coil including the capacity effects of the leads and elements of the tubes.

In case a relatively wide band of frequencies is to be amplified, transformer coupling is usually resorted to, and given suitably designed transformers one to two octaves can be amplified with very fair uniformity at frequencies between 100,000 to 2,000,000 and four to five octaves at frequencies below 100,000 cycles. Use of transformer coupling will broaden the characteristic of the amplifier as shown in curve C, Fig. 28, the exact shape of this curve being largely dependent upon the design of the transformers employed.

For frequencies up to about 100,000 cycles, transformers with iron cores of the ring type are suitable and are preferably enclosed in metal covers which are grounded. A transformer suitable for frequencies higher than 100,000 cycles may consist of two choke coils (one to two inches in diameter) of very fine wire, these coils being mounted close together on a suitable form. The natural frequency of the coils will approximately determine the middle of the band of frequencies which are amplified and the coupling between the two coils will determine the width of the band, closer coupling resulting in a wider band. The coupling is generally a combination of electrostatic and electromagnetic coupling and therefore, in connecting all transformers for high frequency uses, it is essential to establish the proper phase relations between them (see paragraph 21). Each stage of the amplifier should be shielded as shown in Fig. 29 although in certain cases it may be dispensed with. The shielding should

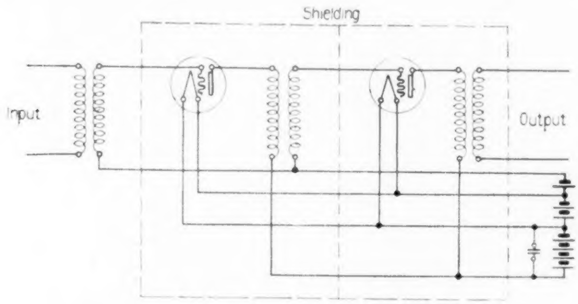


Fig. 29

consist of brass or copper sheeting made into boxes with well-soldered joints and tightly fitting covers. Holes through the shielding should be just large enough to pass the insulation of the wires.

A common plate battery may safely be used for four or more stages provided a condenser is placed across the terminals of the battery as shown in Fig. 29. This condenser should have a capacity large enough to offer practically no impedance to the high frequency currents, and its use may be desirable although the plate battery is common to but two stages. Use of a common grid battery, as shown, introduces a small feed-back from the second stage to the first. This feed-back may be either positive or negative, depending upon the phase relations in the intermediate transformer and may be eliminated by placing a condenser across the grid battery terminals.

It has been the practice in high frequency amplification to use tubes with μ's between 6 and 10, interstage transformers being selected to step up the voltage as much as is possible consistent with the desired flatness of the amplifier characteristic. In general, the larger the ratio of the transformer, the more pronounced is the peak of the characteristic. Other things being equal, the most suitable tubes are those with the smallest internal electrostatic capacities. The largest of these capacities, in general, is that between grid and plate and tubes have been produced in which this does not exceed 5 $\mu\,\mu$ f. and in which the internal plate resistance is about 20,000 ohms.

In amplifying the higher frequencies the feed-back which occurs through the tube may require attention. In section 13, it was pointed out that an inductive output for a tube gives rise to a negative resistance characteristic in the input which means that feed-back is occurring. To eliminate the possibility of singing and also to eliminate unequal amplification of different frequencies which feed-back introduces, various means of neutralizing it have been proposed.[29] One

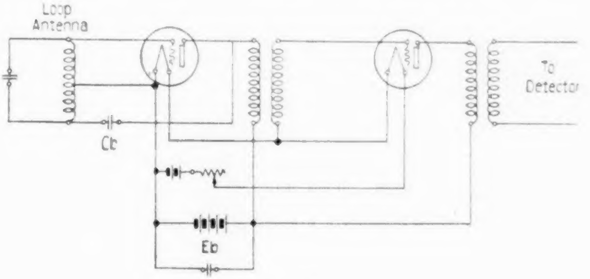


Fig. 30

such means is illustrated in Fig. 30 which is drawn to show radio reception with a loop antenna. Note that the grid of the first tube is joined to one end of the loop and the plate is joined to the other end through the balancing condenser C_b, the filament being joined to the midpoint of the loop. When C_b is chosen equal to C_3 the capacity between grid and plate, it is evident that the feed-back occurring through the tube is just balanced by that occurring through C_b. By adjusting the condenser C_b so as to permit of feed-back, very large amplification may be obtained at a single frequency but at the expense of flatness of characteristic.

[29] See Patent No. 1,183,875 issued to R. V. L. Hartley, and Patent No. 1,334,118 issued to C. W. Rice.

23. *Feed-Back Amplifiers.* This amplifier may be either resistance or inductive coupled, a typical resistance coupled circuit being shown in Fig. 31. In a feed-back circuit, attention must be paid to phase relations. In Fig. 31, let the arrow along the resistance R_1 represent an increase in electron current to the grid of tube A. This corresponds to an increase in the potential of this grid. In phase with this increase in potential is an increase in electron current in R_2 as shown by the arrow. This, in turn, corresponds to a fall in potential of the grid of tube B and therefore to a reduction of the I_p in B, as indicated by the arrow at R_3, which produces an increase in I_p in C. Therefore, in this particular circuit the correct phase relations require the output of one tube to be returned to the input of the second preceding

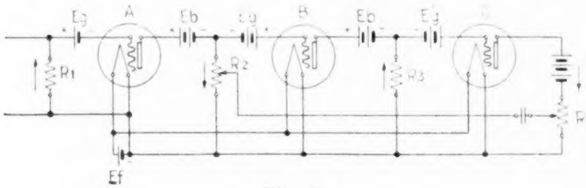


Fig. 31

tube or one of its alternate preceding tubes. The amount of energy fed back can readily be controlled by varying the portion of R_2 through which the feed-back current flows.

24. *Push-Pull Amplifier.* See Fig. 32. This type of circuit is particularly useful as a terminating stage since it makes possible the use of a low impedance in the output circuit without serious distortion. If the tubes A and B have identical characteristics, it is readily seen that the coils of the output transformer may be so connected that the fundamental and odd harmonics will aid one another, while all even harmonics will oppose. Since the third and higher harmonics (counting the fundamental as first) are very small compared to the second, this circuit gives very nearly distortionless amplification. In speech amplifiers it permits of considerable overloading without this being very apparent in the quality of the output.

By reversing the transformer connections it is possible to cause the circuit to add the even harmonics and give the differences of the odd.

An additional use for this circuit will be pointed out in the section dealing with modulation.

Fig. 33 shows a special type of push-pull circuit which is particularly adapted to the amplification of steady and low frequency volt-

ages. It consists of a Wheatstone bridge in which two similar tubes form one pair of arms. The output circuit is the branch in which the galvanometer is ordinarily placed. When a voltage is applied to the two tubes in such manner that the potential of one grid is raised by the same amount as the grid of the other is lowered, the bridge becomes unbalanced and current flows through the output

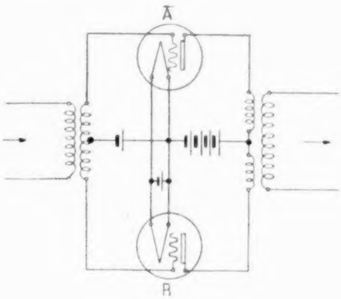


Fig. 32

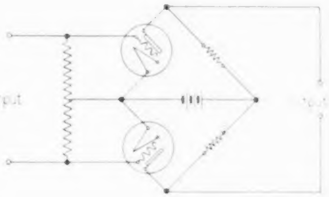


Fig. 33

branch. For small applied voltages the amplification of the circuit is very nearly distortionless. The circuit has the obvious disadvantage of requiring a close balance between the tubes and is therefore liable to require careful adjustment during use.

The push-pull circuit possesses one marked advantage over the resistance-coupled d.c. amplifier described in Sec. 18 for in it current flows through the output branch only when voltage is applied to the input. For the same reason it is also useful for amplifying low frequency alternating voltages.

V. Amplifier Power Supply

The proper power supply for amplifiers is an item of prime importance.

25. *Plate Voltage Supply.* The principal requirement placed on plate voltage is that it be steady. For this reason storage batteries are usually best, but good dry cells are more often used and when fresh, prove very satisfactory. The principal trouble encountered in the use of dry cells arises from an attempt to use partially rundown cells. A dry battery should be tested periodically for voltage, the reading being taken while the battery is delivering a current at least as large as that drawn by the amplifier. Whether dry cells or storage cells are used for plate voltage, in general not more than four stages

should be operated from a single battery. In the case of a retard coupled amplifier whose stages are tuned with condensers, each stage should preferably have a separate plate battery to reduce the tendency to "sing."

A generator as a source of plate voltage is frequently used for power amplifiers. In case a direct current generator is used, a filter is generally necessary in the plate circuit to remove commutator ripples.

26. *Filament Voltage Supply.* A source of constant filament voltage is not necessary in order to insure constant space current within the tubes at temperature saturation, but in general any variation in filament current will affect the relative potential difference between filament and grid, and is, therefore, equivalent to a variation in the input voltage. This possible source of trouble must be particularly guarded against in such a circuit as that shown in Fig. 25, in which a portion of the adjustable resistance of the filament circuit is included between the filament and the "common point" A. If storage batteries are available, they form the best source of filament current; generators have been used satisfactorily however.

27. *Sources of Grid Potential.* A flow of electrons to the grid of a tube is liable to result either in distortion or a loss of amplification or both (see Secs. 15 and 16). The steady negative grid potential required to prevent the input voltage carrying the grid to a positive potential may be obtained from either one of two sources: by a grid battery or by an *IR* drop in some resistance in the circuit. The requirements for a grid battery are very light since it is called upon to give no appreciable current. As was pointed out in Sec. 26, use of an *IR* drop for grid voltage pre-supposes steady filament or plate battery according to circumstances, and since the proper grid battery is readily obtainable the use of an *IR* drop is likely to prove desirable only in very unusual circumstances.

VI. Troubles in Amplifier Circuits

28. *Noise.* The noise in amplifier circuits is due to several causes which may, in general, be grouped into two classes. Certain noises originate within the tubes and other noises find their origin in the circuit. The amount of noise in any amplifier limits the minimum input voltages which it will handle satisfactorily, for obviously input voltages which produce output currents of the same order of magnitude as the currents giving rise to noise will not be satisfactorily amplified.

29. *Tube Noises.* Tubes may be responsible for three distinct kinds of noises. (a) Ringing or rattling is due to the vibration of the tube elements and may be eliminated by proper tube construction or by some form of vibration proof suspension for the early stages of the amplifier. (b) Crackling may be produced by high resistance films on the inner surface of the bulb, forming conducting paths between the leads. Faulty electrical contact between the plate, grid and filament and their respective leads is also a frequent source of crackling. Furthermore, in tubes which are well constructed in regard to the points just mentioned, but which contain tungsten filament, crackling may be observed. This trouble is not found in all tungsten filament tubes but, when present, is sufficiently marked to become apparent in a two stage amplifier. (c) In carefully constructed amplifiers of more than three or four stages a noise which can best be described as a hissing or sighing is certain to be present. It appears to be related to an unavoidable statistical variation in the escape of electrons through the grid to the plate. Its magnitude has been found to correspond approximately to an output voltage from the first stage of between 5×10^{-7} volts and 5×10^{-6} volts. Between these limits the noise is found to increase as the output impedance of the first stage is increased, and also to increase as the resistance across the terminals of the input increases. Its components, above 300 cycles, appear to be of about equal magnitude and uniformly distributed. It is, therefore, impossible at the present time to build amplifiers to handle voltages of less than this order of magnitude, at any rate when the frequencies involved are in the audible range.

30. *Circuit Noises.* In general, circuit noises in amplifiers are due to one or more of the following causes: variations in grid and plate batteries, loose contacts and variations in resistances, leakage of condensers and leakage across the insulating mounting upon which the amplifier parts are fastened, and external electric or magnetic fields acting inductively on the circuit. The remedy in each case is obvious once the exact cause has been found. To eliminate inductive effects in the wiring it is usually sufficient to run wires in pairs and to shield them electrically, the shielding being grounded. In laying out the various parts of an amplifier it is well to place the bulky pieces at points in the circuit at which they will have as near zero potential as possible.

31. *Singing.* Singing, which is one of the most serious troubles in amplifiers, is always due to some form of feed-back. This may be magnetic, electrostatic, or in the form of mechanical vibrations as in an amplifier having a microphone attached to the input and a receiver

to the output. Mechanical feed-back can also occur in the case of tubes whose parts can easily be set into vibration and a cure is usually found in some form of vibration-proof mounting. The coupling which is responsible for feed-back may be difficult to locate, but when found can usually be removed. Both retard coils and transformers may afford an easy method of coupling due to stray fields. If the coupling induces voltages which are in phase with the input voltages, it may cause singing, and if out of phase, the amplification may be seriously reduced. Closed core coils and magnetic shields are the usual remedies for this condition, although a rearrangement of the circuit parts may be necessary.

Certain kinds of electrostatic feed-back may be removed by enclosing each stage in a separate grounded metal cage or box. The electrostatic coupling due to tube capacities (Sec. 13) cannot be eliminated but it is possible to so design circuits that trouble from this source will not present itself. Thus an inductive impedance in the output circuit may prove troublesome because it induces a *negative* resistance back in the input circuit; a non-inductive output can never do this. Feed-back through tubes increases with frequency, and in the case of high frequencies, it may sometimes be necessary to use resistance coupled rather than reactance coupled circuits.

32. *Blocking.* Two entirely different types of blocking may occur in an amplifier. They both result from the grid of one or more tubes having been carried to a positive potential by the input voltage. While positive, the grid picks up a negative charge of electrons which is removed more or less rapidly by the grid leak. In case the leak resistance is high, a residual charge may remain upon the grid for an appreciable length of time, depressing its mean potential to so low a value that the output of the tube is cut to zero or very nearly zero. The remedy is obviously to reduce the input voltage or to increase the voltage of the negative grid battery. In certain cases, a readjustment of the resistance of the grid leak may be desirable.

The second type of blocking involves secondary emission from the grid as discussed in Sec. 9. It can occur only when the input is sufficient to force the grid potential of some tube positive by as much as 10 or 15 volts, and then if the grid leak resistance is large enough, secondary emission will hold the grid at about this positive potential and entirely prevent proper functioning of the amplifier. In eliminating this type of blocking, the first step should be to note the effect of increasing the filament currents as secondary emission is less likely to occur when the filament yields a copious supply of electrons. If this does not remove the trouble, the negative grid batteries in the

stages at fault may be increased and lower grid leaks may be desirable. The volume of input to each stage should also be considered.

33. *Distortion.* Distortion in an amplifier circuit may result either from a failure to amplify all frequencies by the same amount or from the generation of overtones of the fundamental frequencies in the input.

The unequal amplification of various frequencies arises from the presence of resonant characteristics in the circuit. This may take the form of a feed-back which discriminates in favor of certain frequencies, the feed-back not being pronounced enough to cause singing. A negative feed-back may also occur, causing a loss of efficiency over some particular frequency range.

The distortion which arises from the generation of overtones is due to non-linear voltage-current characteristics in one or more branches of the circuit. The usual sources of this trouble are curvature of the plate and grid characteristics (See Equation 4) and the variable permeability of the iron used as cores. With properly chosen coils, practically distortionless amplification can be secured by the method indicated in Sec. 12. In general, to accomplish this, the output impedance need not be more than two or three times r_p. In case it is necessary to use a low output impedance in the final stage, distortion may be reduced by using the push-pull circuit of Sec. 23.

In using an amplifier under circumstances such that distortionless output is desired, care should be taken that no tube by itself is overloaded or caused to work in such fashion that its dynamic characteristic is curved. Distortion which arises from curvature of this characteristic can be detected by inserting an ammeter in the plate circuit of each tube. When each characteristic is straight, or nearly so, there should be no change in ammeter reading as the source of input voltage is thrown on and off, and in the case of a variable input such as that arising from speech, the ammeter readings should remain constant while the amplifier is in operation. This test will not detect distortion which arises from selective amplification with respect to frequency.

34. *Calculation and Measurement of Amplification.* Provided all parts of an amplifier circuit are functioning properly and its constants are known, its amplification can be calculated quite accurately. The following example will illustrate the procedure to be followed in any case. Referring to the transformer coupled amplifier of Fig. 26, assume that the ratio of the first input transformer is $z:z_1$ and that the ratio of the second input transformer is $z_2:z_3$; assume also that z_2 is numerically equal to r_p, the plate circuit resistance of the first tube.

Then, calling e_g the input voltage, the voltage across the first tube is

$$e_g\sqrt{\frac{z_1}{z}};$$

the voltage across the primary of the second input transformer is

$$e_g\frac{\mu}{2}\sqrt{\frac{z_1}{z}}$$

since z_2 is numerically equal to r_p; and across the secondary is

$$e_g\frac{\mu}{2}\sqrt{\frac{z_1}{z}}\sqrt{\frac{z_3}{z_2}}.$$

Hence the voltage amplification of this portion of the amplifier is

$$\frac{\mu}{2}\sqrt{\frac{z_1}{z}}\sqrt{\frac{z_3}{z_2}},$$

and a similar argument applies to the following stages.

The measurement of amplification can be accomplished by the obvious procedure of determining the magnitudes or the relative magnitudes of the input and output current. This can be done for either a single stage or for several stages at once.

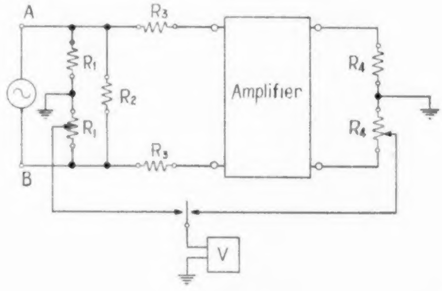


Fig. 34

A very satisfactory circuit for measuring amplification is illustrated in Fig. 34 and through use of a vacuum tube voltmeter (see Sec. 56) as a comparison means, it is capable of an accuracy of about 2%. The conditions which the resistances R_1 and R_2, etc., should satisfy are very simple. The network connected between the oscillator and amplifier input should present an impedance, looking into it from the right, equal to the input impedance of the amplifier; like-

wise the output impedance of the amplifier should equal $2R_4$. The input impedance of the vacuum tube voltmeter is so high as not to shunt the resistances across which it is connected appreciably. The grounds at the mid points eliminate the disturbing effects of capacities to ground. Under these conditions the voltage amplification a is given by the equation:

$$a=\frac{\alpha}{\beta}\left[\frac{4R_3(2R_1+R_2)+2R_1R_2}{2R_3(2R_1+R_2)+2R_1R_2}\right],$$

in which α, β are the fractions of R_1 and R_4 respectively, across which the voltmeter is connected to obtain equal readings when the switch W is thrown from one position to the other. In case R_2 is made quite small with respect to R_1, R_3, the expression for a reduces approximately to $a=\frac{2\alpha}{\beta}$. An expression for current amplification can readily be derived.

Another simple measuring circuit is shown in Fig. 34*a* in which O is an oscillator of the desired frequency, F is a filter to remove harmonics from the oscillator current. R_1R_2 and R_3R_4 are attenuating networks consisting of resistances, WWW are switches by which the telephone receiver T can be joined either to the output of the amplifier

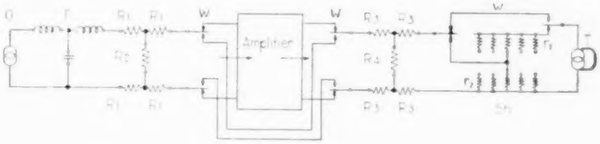


Fig. 34a

or directly to the oscillator, and also provide means for removing the known attenuation in the shunt *Sh* (receiver shunt) at the same time the amplifier is removed. By the proper design of the receiver shunt, which will be discussed presently, the attenuation required to give the same volume of sound in the receiver whether the amplifier is in or out may be read directly.

In setting up the circuit of Fig. 34*a*, special attention must be given to the networks R_1R_2 and R_3R_4. In addition to reducing the input to the amplifier to a value safely below the overload point, R_1R_2 should be designed to present an impedance (when seen from the amplifier) equal to that out of which the amplifier is to operate in service. Otherwise the measurements of amplification may be without significance.

The network R_3R_4 serves two important purposes. It is designed to present toward the amplifier the same impedance as the amplifier is to work into in service, and this in turn requires that the input and output impedances of the amplifier be practically equal (or if not, then small with respect to R_1) for otherwise the network R_3R_4 when joined to R_1R_2 will not draw the same fraction of current as the amplifier, thereby upsetting the comparison upon which the measurements are based. Furthermore, the attenuation in R_3R_4 is to be sufficiently large that variations in the impedance of the receiver and its shunt as seen from R_3R_4 will not appreciably affect the impedance into which the amplifier works as the receiver shunt setting is changed. A simple calculation will show how great the attenuation must be in any given case to satisfy these conditions.

Proper values for the steps of the receiver shunt may be calculated as follows, reference being made to Fig. 35 in which the currents and potentials indicated are in accordance with the assumptions made

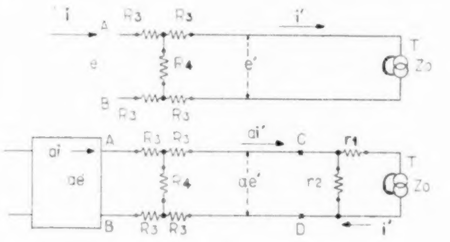


Fig. 35

regarding the attenuation in the various portions of the circuit. Calling a the amplification to be measured it may readily be shown that

$$\frac{(a-1)^2}{a}=\frac{r_1}{r_2}. \tag{9}$$

Or if R is the impedance of the network R_3R_4 as seen from the receiver, and the shunt Sh is proportioned so that it also presents the impedance R to the receiver, whence

$$R=r_1+\frac{r_2 R}{r_2+R},$$

then Equation (9) gives

$$a=\frac{R+r_2}{r_2}.$$

Taking account of the necessary approximations it is readily possible to measure current amplification to within 5%, for a range of

frequencies extending from 200 to 3,000 cycles. Receiver shunts are made which, in 10 to 15 steps, will reach a maximum reduction ratio in current of 25:1 which corresponds to an energy reduction of 625:1, and this does not represent the greatest range possible.

In case a rougher approximation of the amplifying power is sufficient, the circuit of Fig. 34*a* may be simplified by omission of the network R_3R_4 and reversal of the receiver shunt to present a constant impedance (except for variation of impedance with frequency and phase angle) toward the amplifier. The network R_1R_2 should preferably be retained and should be so proportioned that the current through the right hand R_1 branches is practically the same whether connected with the amplifier or directly to the receiver.

In measuring the over-all amplification of a multistage circuit, it will probably be desirable to add fixed but known attenuation units similar to R_3R_4 to the receiver shunt which may be cut in or out as required. These units may be given an attenuation equal to and twice the total attenuation of the shunt, etc., after the fashion of the ordinary resistance box. In constructing attenuation networks the arrangement indicated in Fig. 34*a* will be found desirable in that the symmetrical placing of the branches tends materially to eliminate errors which might otherwise arise due to capacities to ground in the oscillator and amplifier. Pairing of lead wires and shielding of leads and resistance coils will be found desirable for accurate work.

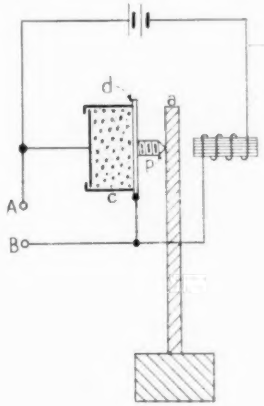


Fig. 36

The filter *F* should be used in case the amplifier tends, because of the limited range of frequencies which it passes or because of some other kind of distortion, to modify the quality of the note given by

the oscillator, it being very difficult to match sounds for intensity which differ in quality.

A very satisfactory type of audio-frequency generator is shown in Fig. 36; it is a buzzer which operates, not by making and breaking current, but by varying it periodically with a microphonic button. The vibrating parts of this generator may be tuned to any audio-frequency, e.g. 800 cycles, and it gives quite accurately a sinusoidal variation of current, although it is customary to insert a filter (Sec. 6) to insure the input energy being accurately of one frequency.

VII. Thermionic Modulators

In discussing modulation the terminology which has been developed in connection with radio and carrier-current signaling will be used.

By the term "modulation" is meant the varying of the amplitude of a relatively high frequency wave, so that its envelope represents a particular low frequency wave or combination of such waves. (See Curves A, B and C, Fig. 37). The combination of low frequency

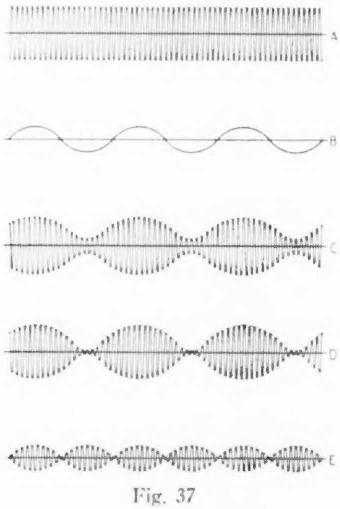


Fig. 37

modulating waves may be very complicated, as in the case of speech, but the principle involved is common to all cases of modulation and can be clearly brought out by the consideration of a single low frequency.

Let Fig. 37, C, represent a high frequency wave modulated by a sinusoidal low frequency. The wave C can be represented by

$$y = a(b + \cos qt)\cos pt, \qquad (10)$$

in which $\frac{p}{2\pi}$ is the high (carrier) frequency and $\frac{q}{2\pi}$ the low (signal) frequency. Equation 8 can be rewritten in the form

$$y = ab\cos pt + \frac{a}{2}[\cos(p-q)t + \cos(p+q)t], \qquad (11)$$

which brings out the fact that the modulated wave C contains, in general, three distinct frequencies—the carrier frequency $\frac{p}{2\pi}$, a difference frequency $\frac{p-q}{2\pi}$, and a summation frequency $\frac{p+q}{2\pi}$. These latter frequencies represent the so-called "side bands" of the modulated wave.

Two special cases of the wave represented by Equations 10 and 11 are represented graphically at D and E of Fig. 37 and correspond to $b=1$ and $b=o$ respectively. When $b=1$ it is evident that the amplitude of each side band is half the amplitude of the carrier frequency; such a wave is said to be "completely modulated"; when $b=0$ the carrier frequency $\frac{p}{2\pi}$ is absent altogether.[31]

35. *Means for Producing Modulation.* Perhaps the simplest case of modulation is that illustrated by continuous-wave radio telegraphy, in which the intermittent radiation of a uniform wave is accomplished by means of a telegraph key. In most cases, however, modulation requires a gradual change in the amplitude of the high frequency wave. For effecting this the vacuum tube possesses two properties which make it particularly useful—(a) the E_g, I_p characteristic is very nearly parabolic (Sec. 8); (b) the current in the plate circuit is a function of the grid potential (Fig. 10).

Circuits, by means of which modulation may be effected by each of these properties, are described in the following paragraphs.

36. *Modulation by Curved Characteristic.* Considering the circuit of the type illustrated in Fig. 38, let it be assumed that a voltage

$$e = A\cos pt + B\cos qt$$

is applied to the input of the tube. The result is shown graphically in Fig. 39. When this value of e is substituted in Equation 4 we

[31] For a more complete discussion of modulation and the nature of the side bands, see R. V. L. Hartley, Procd. Inst. of Radio Engrs., Feb., 1923, or *Bell System Technical Journal*, Apr., 1923.

obtain for the modulated output (i.e., terms whose frequencies are of the order $\frac{p}{2\pi}$).

$$J_m = A\left[\frac{\mu}{r+r_p} + \frac{\mu^2 r_p r_p'}{(r+r_p)^3} B \cos qt\right]\cos pt. \tag{12}$$

As pointed out above, the first term gives the carrier wave and the second term, the two side bands. It will be remembered that this

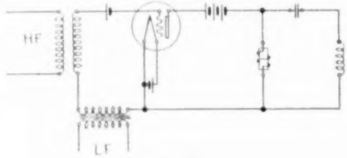


Fig. 38

equation neglects terms of higher order than the second, which is permissible, so long as the tube characteristic is approximately parabolic.

Certain points regarding Equation 12 should be noted. In the first place, the amplitude of the side bands is proportional to the

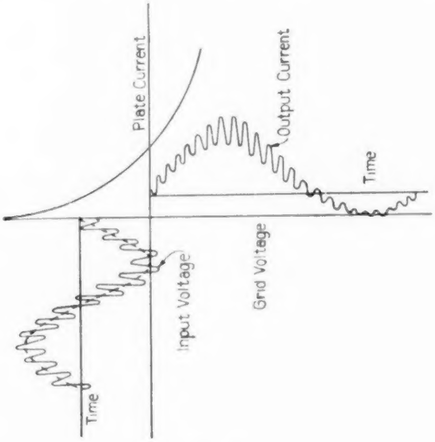


Fig. 39

product AB and is therefore, independent of the relative amplitudes of the original carrier and modulating frequencies. Also the modulated current is proportional to the first power of B, the amplitude

of the low frequency wave; i.e., although the modulation is effected by the curvature of the tube characteristic, the modulated output is free from distortion. Furthermore, the modulated output voltage is proportional to $\frac{r}{(r+r_p)^3}$ which is a maximum when $r=\frac{1}{2}\,r_p$; and the modulated output energy is proportional to $\frac{r}{(r+r_p)^6}$ which is a maximum when $r=\frac{1}{5}\,r_p$.

Another type of circuit in which the modulation is dependent upon the curvature of the E_g, I_p characteristic is shown in Fig. 40. The two tubes are supposed to be alike; so long as no low frequency is impressed on the grids the high frequency space currents are equal,

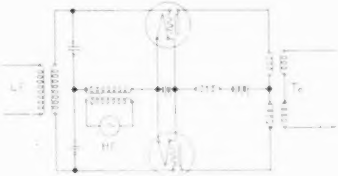


Fig. 40

each passing through one of the primary coils of the transformer T_0, the order of winding these coils being such that this condition gives zero current in the secondary. However, the presence of a low frequency voltage (*L.F.*, Fig. 40) raises one grid potential at the same time that it lowers the other, with the result that the high frequency currents in the two primary coils are no longer equal, and a high frequency current therefore flows in the secondary of T_0, the amplitude of which is determined by the degree to which the two tubes are unbalanced by the low frequency input. It is apparent that the output of this modulator circuit contains only the two side bands and none (or very little, if the tubes are not exactly identical) of the carrier frequency and therefore corresponds to curve *E* in Fig. 37. It is particularly useful in communication circuits where several telephone or telegraph channels are desired on the same pair of wires. Since only the side bands are transmitted the total current which must be handled by repeaters and other line apparatus is materially reduced. By the use of the proper wave-filter it is also possible to suppress one side band, thereby approximately cutting to one half, the width of the frequency band to be transmitted. As will be pointed out under homodyne detection, the suppressed carrier frequency must be supplied locally before detection can occur.

37. *Modulation Effected by Controlling Plate Current with Grid Potential.* Numerous circuits have been developed for modulation, making use of the fact that the grid potential affects the resistance of the plate circuit. Two circuits of this type are shown in Figs.

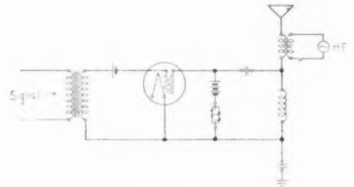


Fig. 41

41 and 42. In the first, the plate circuit of the tube forms a shunt across a portion of the antenna inductance. As the grid potential is varied, the antenna is, therefore, thrown more or less out of tune, with the consequent radiation of a variable amount of high frequency energy.

Fig. 42 shows one of the most efficient modulating schemes thus far developed. As the grid potential of the modulator tube *A* varies, causing a change in plate current through this tube, the plate voltage

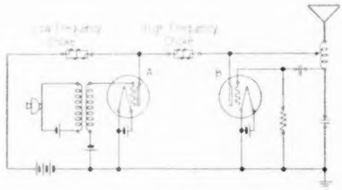


Fig. 42

applied to the oscillator tube, *B*, fluctuates, because of the presence of the low frequency choke coil. Under this condition of variable plate-voltage the oscillator gives a variable amount of high frequency energy to the antenna. When used for speech modulation this circuit is very efficient and gives good quality. The tubes *A* and *B* are ordinarily of the same type.

Many other modulating circuits have been designed, and those given above are to be considered merely as illustrative of the general manner in which the properties of the vacuum tube may be applied to the problem of modulation.[32]

[32] For other types of modulator circuits see a paper by R. A. Heising, Procd. Inst. Radio Engrs., Aug., 1921.

VIII. Thermionic Detectors

Like the modulator the detector is a device for the production and separation of difference frequencies. The object of modulation is, in general, to transform a high frequency $\frac{p}{2\pi}$ and a low frequency $\frac{q}{2\pi}$ into two high frequency side bands, $\frac{p \pm q}{2\pi}$. Detection accomplishes the inverse operation of forming from a carrier frequency $\frac{p}{2\pi}$ and either or both side bands the original low frequency $\frac{q}{2\pi}$, detection often being referred to as demodulation. Detection, like modulation, can be most readily described by the consideration of a single pair of frequencies.

When carried out by means of a vacuum tube it results from rectification in either the grid circuit or the plate circuit. This rectification

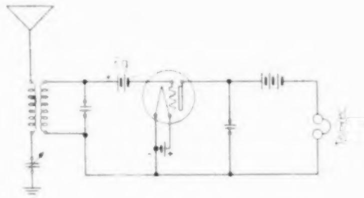


Fig. 43

may arise either from unilateral conductivity or a curved current-voltage characteristic as pointed out in the following paragraphs.

38. *Detection by Curved Plate Characteristic.* Considering the circuit shown in Fig. 43 and assuming an input voltage $e = A(B + \cos qt)\cos pt$, it follows from Equation 4 that the output current, considering only those terms whose frequencies are of the order $\frac{q}{2\pi}$, is

$$J_d = \frac{1}{2!}\frac{\mu^2 r_p r_p'}{(r+r_p)^3} A^2 \left(B \cos qt + \frac{1}{4}\cos 2\,qt\right). \qquad (13)$$

The current J_d, known as the "detected current," therefore, consists of a term whose frequency is $\frac{q}{2\pi}$ and another term whose frequency is twice this. The presence of these two frequencies is readily understood. The detected current of frequency $\frac{q}{2\pi}$ corresponds to the

difference frequency (See Equations 10 and 11) of the carrier of amplitude AB and each side band of amplitude $\frac{A}{2}$ and is therefore proportional to $2 \cdot \frac{A}{2} \cdot AB$. The second term of the detected current represents the difference frequency $\frac{q}{\pi}$ between the two side bands themselves and, as is to be expected, its amplitude is proportional to $\frac{A^2}{4}$. In case one of the side bands is suppressed before detection, this term of double frequency is entirely absent in the detected current. Furthermore, the amplitude of the detected current of frequency $\frac{q}{2\pi}$ is independent of the *relative* amplitudes of the carrier wave and the side bands. In general, AB is large compared to $\frac{A}{2}$ with the result that the term of double frequency in the detecting current is negligible. It follows, therefore, as in the case of modulation that the detecting current is practically free from distortion.

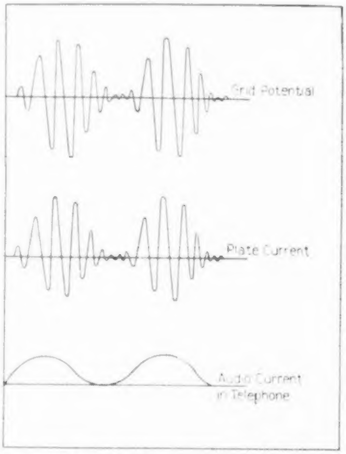


Fig. 44

The detecting action resulting from the curved plate characteristic is shown in Fig. 44.

Equation 13 leads to the result that the output voltage of a detector tube, when working as above, is a maximum when $r = \frac{1}{5}\, r_p$. In using these relations note that r represents the value of the output resistance

for the high frequency and not the low frequency. When using an amplifier on the output of a detector (see Fig. 45), it is important to choose r to give the maximum detecting *voltage*. As in amplifier cir-

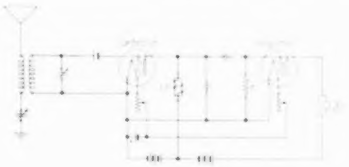


Fig. 45

cuits it is essential in this type of detector that the grid remain always negative. See also the sections on amplifiers.

39. *Detection by Rectification in Grid Circuit.* This type of circuit (see Fig. 46) is now in very general use for radio purposes, and is characterized by the grid blocking condenser C_s. Contrary to the preceding type of detector, the present requires the flow of electrons to the grid and works best when the grid is held permanently at a

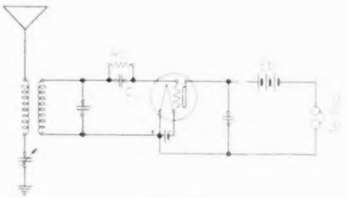


Fig. 46

small positive potential. The action of the high frequency input causes the periodic accumulation of a negative charge upon the grid and the blocking condenser, thus lowering E_g and diminishing I_p. This action is clearly illustrated in Fig. 47. This circuit is most effective when the carrier frequency is much greater than the signal frequency, and not as efficient as the circuit described in Sec. 38 when the carrier is say only four or five times as great as the signal frequency.

Attempts to deduce a quantitative relation for the detecting current in this type of detector have as yet met with little success, one of the principal reasons being that very little is known about the "dynamic" grid current characteristic.[33] Experiments show, however, that the

[33] For a discussion of this topic see Hulbert & Breit, *Phys. Rev.*, Nov., 1920, pp. 408–419; Oct., 1920, pp. 274–281.

detecting current is practically proportional to the square of the input voltage, provided that this is small, thus establishing the relation,

$$J_d = ae^2,$$

which corresponds in form to Equation 13 above.

In designing this type of detector circuit, attention must be paid to the value of the blocking condenser C_s and its leak R_s. It is clear that the capacity of C_s should be sufficiently small to cause the grid to undergo the maximum potential change as a result of the relatively

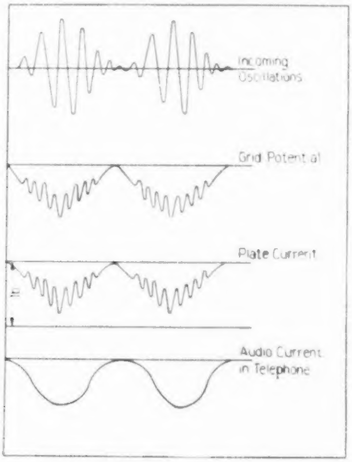


Fig. 47

small electron charge picked up, and yet it must be several times larger than the tube capacity between grid and filament. Furthermore, the time constant of C_s and R_s should approximately match the frequency of the detected current. With the more common detector tubes and radio frequencies, capacities of the order of 200 $\mu\mu$f. are satisfactory.

Detector circuits with grid blocking condensers may be coupled to amplifiers as readily as the other type of detector, and, in general, a higher output resistance for the detector can be used, thus making possible more efficient coupling between the detector and the first stage of amplification. In increasing the output resistance of the detector it should, however, be borne in mind (see Sec. 13) that a secondary result is to reduce the input impedance of the detector, which may entail a reduced input voltage.

40. *Heterodyne and Homodyne Detection.* In continuous-wave radio telegraphy, the dots and dashes of the code are transmitted by a continuous carrier wave of a single frequency. *Heterodyne* reception consists in supplying a slightly different frequency at the receiving station, the transmitted and locally generated frequencies when applied to the detector acting exactly as the carrier and side band frequency described above. The useful output of the detector is the difference frequency which, of course, is chosen in the audible range.

It follows from Sec. 38 that the heterodyne detecting current is proportional to the product of the amplitudes of the transmitted and locally generated waves. Because of this fact a feedback circuit may be used to advantage as a means of increasing the strength of both high frequency terms. In the usual type of feed-back detector, the detector tube is also used as the source of local high frequency. Such

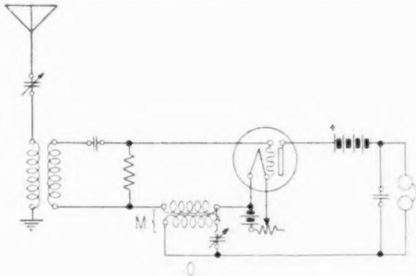


Fig. 48

a circuit is shown in Fig. 48. The oscillatory circuit *O* is tuned to differ in frequency from the incoming signal by an amount which will give a satisfactory difference frequency in the telephone receiver *T*. By varying the coupling at *M*, the intensity of the beat note can be readily changed. It must, however, be sufficient to cause the circuit as a whole to oscillate at the natural frequency of *O*. A feed-back arrangement is particularly applicable to those cases in which the carrier frequency is much higher than the signal frequency and is therefore generally used with a blocking condenser.

In telephone systems, whether radio or carrier current, it is frequently desirable to suppress the carrier frequency and transmit only one[34] or both of the side bands. Detection with only the side bands present would result in a double frequency detecting current which obviously would not be permissible in a telephone circuit. Whenever

[34] For a discussion of the advantages of *single* side band transmission, see reference given in footnote 13.

the side bands alone are transmitted, a locally generated high frequency exactly equal in frequency and phase to the original carrier frequency must be supplied.[35] This is known as *homodyne* reception.

41. *Measurement of Detection Coefficient.* The constant a in the relation $J_d = ae^2$ is called the "detection coefficient." Its measurement by direct means is not difficult but as seen from Equation 13 it involves so many factors that no satisfactory indirect methods of determination have been developed. The requirements of the direct method are quite obvious, and for circuit details reference is made to the *Thermionic Vacuum Tube* by van der Bijl.

42. *Detecting Efficiency.* A knowledge of the detecting coefficient a tells very little about the detecting efficiency of a tube, the efficiency being defined as the ratio between the low frequency energy in the output and the high frequency energy in the input. The efficiency involves the input impedance of the tube which is a function of the circuit constants as well as the tube. It is therefore impossible to specify the detecting efficiency of a tube without certain data concerning the circuit in which it is to be used; a is therefore without much significance.

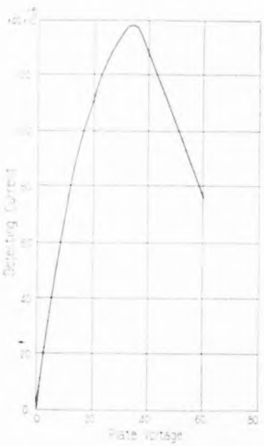


Fig. 49

43. *Detecting Coefficient and Plate Voltage.* The variation of the detecting coefficient with plate voltage depends upon the type of circuit. If detection is accomplished by a curved plate characteristic, experiment shows (see Fig. 49) that the operation is best when the effective voltage is about equal to the potential drop in the filament,

[35] See J. R. Carson, Procd. Institute of Radio Engineers, Vol. II, p. 271, 1923.

it being presupposed (see Sec. 38) that E_g is enough less than zero to keep the grid negative at all times. If detection is accomplished by means of a blocking condenser, the variation is as shown in Fig. 50, no sharply defined maximum being present in the curve.

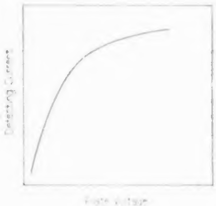


Fig. 50

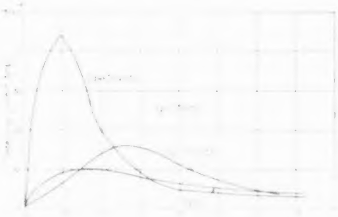


Fig. 51

The variations of detected current under heterodyne operation are shown in Figs. 51 and 52 which refer respectively to detection with and without a blocking condenser. The abscissa, e_1e_2, gives the product of the two high frequency amplitudes in the input. As is to be expected when no grid condenser is used (see Equation 13), the

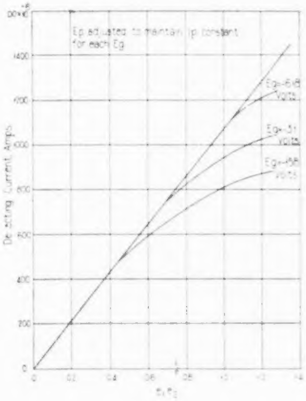


Fig. 52

variation of detected current with e_1e_2 is very nearly linear provided E_g is sufficiently negative. Fig. 51 referring to detection with grid condenser shows no linear relation however. The data for Figs. 51 and 52 are taken from Van der Bijl.

44. *Comparison of Tubes as Detectors.* If a tube is available whose detecting coefficient is known, other tubes may be calibrated in terms

of this standard. The comparison of detectors can be very readily carried out by means of such a circuit as shown in Fig. 53. This circuit makes use of a grid blocking condenser but could readily be rearranged not to employ it. In use, switches S and K are operated together in such manner that the receiver shunt is cut out when the

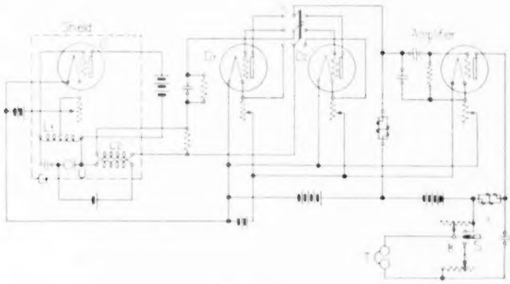


Fig. 53

receiver is connected to the tube of lower detecting power. By proper adjustment of the shunt the two tubes can then be brought to apparent equality, the difference being read from the calibration of the shunt.

IX. Vacuum Tube Oscillators

As pointed out in the section on amplifiers, it is easy to design feed-back circuits which will sing, i.e., will generate continuous oscillations. The necessary requirements which an oscillating circuit must meet are two in number and are readily understood. Any small alternating voltage when applied to the input generates a current in the output, and by virtue of the feed-back a portion of this energy is returned to the input. For continuous oscillations, the energy returned must be in phase with the original input supply. Furthermore, letting e represent the initial input voltage, the feed-back coupling must be sufficient to return to the input a voltage greater than e. If it is less than e the circuit will amplify but will not oscillate.

The circuit requirements necessary for any given tube to return a voltage greater than e may readily be stated in mathematical form, but so far as the practical design of oscillators is concerned, this statement has no particular value. The design of circuits is still very largely an empirical matter, and the problem is not so much to make the circuit oscillate as to make it oscillate with the proper frequency, efficiency and output power. These requirements can usually best be

met by trial and adjustment taking into account such general theoretical considerations as follows.

Vacuum tube oscillators make a convenient way of obtaining large high frequency currents at small voltages and large a.c. voltages at small currents for testing purposes.[36] A special oscillator circuit for giving a very pure sine wave output of constant frequency is discussed in Sec. 53.

45. *Equivalent Resistance of the Oscillator Circuit.* Oscillator circuits are of many types, but the fundamental action of all of them can be reduced to common terms, and to simplify the discussion the type of circuit illustrated in Fig. 54 will be discussed.

It will be noted that a d.c. voltage is applied between filament and plate by means of a battery in series with a choke coil. This choke

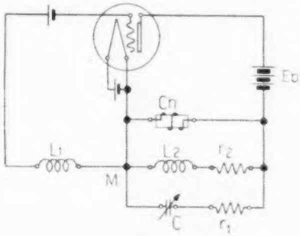


Fig. 54

may be considered as having zero d.c. resistance and virtually an infinite resistance to the a.c. generated by the oscillator. An oscillating circuit consisting of inductance, resistance and capacity is also connected between filament and plate, and by means of another inductance joining filament and grid a portion of the output energy is fed back to the input by virtue of the inductive coupling M.

In circuits as usually constructed the ohmic resistances r_1 and r_2 of the oscillating circuit are very small compared to the impedances of L_2 and C. It follows that the frequency of oscillation differs but slightly from the natural frequency $1/2\pi\sqrt{L_2C}$, and the oscillating circuit may therefore, be looked upon as introducing nothing more than a pure resistance (so far as the fundamental component of I_p is concerned) into the output circuit of the tube. Except for considerations of feed-back, we may therefore imagine the oscillating circuit L_2, C, replaced by an equivalent resistance whose value is given by the equation $R=\dfrac{r_1r_2+L_2/C}{r_1+r_2}$. This "equivalent resistance" is important in dealing with oscillating circuits.

[36] See W. C. White, Gen. Elect. Rev., Vol. 20, p. 635, 1917.

46. *Phase Relations.* Typical phase relations between the various currents and voltages in the oscillating circuit are illustrated by the oscillograms in Fig. 55. Note that the oscillation current I_o, and also E_p and E_g show practically sinusoidal variations. Such variations

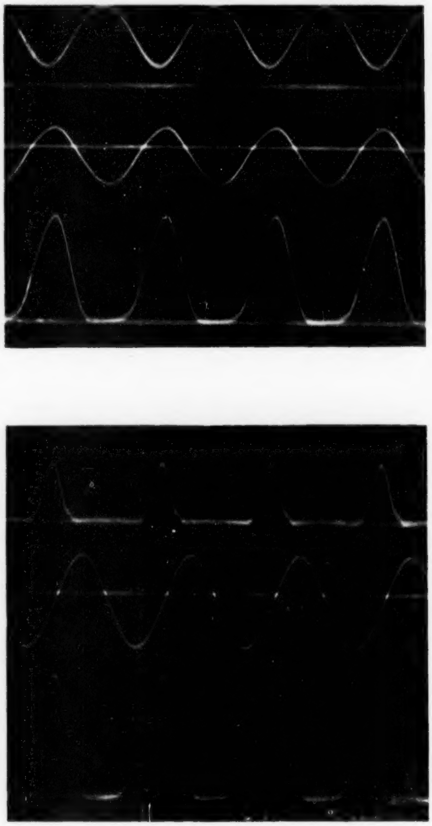


Fig. 55

will be found over very wide ranges of adjustment. Although I_p does not show a sinusoidal variation, the fundamental component (whose frequency is the same as that of I_o) is relatively much larger than the higher components.

Note that the variations of E_p and E_g are practically 180° out of phase, and also that E_p and the fundamental of I_p are 180° out of

phase. This latter condition is obviously required for maximum power output by an a.c. generator. In order that I_p be a maximum when E_p is a minimum, E_g must be a maximum at this time.

47. *Dynamic Characteristic.* A tube when operating into an output resistance follows a dynamic characteristic (also called derived characteristic)[37] whose slope is somewhat less than that of the static (see Sec. 12). The dynamic characteristic is flatter than the static characteristic since the oscillating circuit L_2, C acts as an equivalent resistance. In fact, the dynamic characteristic of the oscillating tube may differ in one important respect from that of Sec. 12, for in an oscillator the grid potential has a relatively very high positive value for a portion of the cycle. Consequently the dynamic plate characteristic very frequently turns downward at its upper end (point B, Fig. 56). As will be pointed out later, this feature of the dynamic

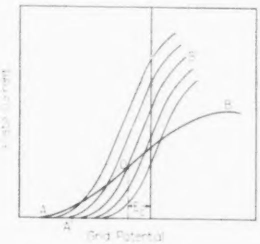


Fig. 56

characteristic is likely to be one of the factors tending to limit the amplitude of oscillation.

48. *Amplitude of Oscillation.* As yet no very comprehensive formula has been derived, either theoretically or empirically, to express the amplitude of oscillation in terms of the tube and circuit constants and the applied E_p. However, certain general statements can be made which will serve as useful guides.

We shall not consider further the circuit shown in Fig. 54; it shows a negative grid battery in the grid filament branch, which, if sufficiently large to control the oscillator when in operation, usually makes the starting difficult and uncertain, and is therefore undesirable. To eliminate the grid battery and yet supply sufficient negative potential, once oscillations have started, circuits are usually supplied with a grid blocking condenser and high resistance leak (see Figs. 64 and 65). Since for a portion of each oscillation the grid is positive, it

[37] See L. A. Hazeltine, Procd. Inst. of Radio Engrs., April, 1918.

picks up a charge of electrons which in flowing off through the leak creates an average negative grid potential. It is apparent that as the amplitude of oscillation increases, the charge picked up at each positive swing of the grid potential increases, with the result that this average negative potential tends to sink lower. The importance of this control feature will be brought out presently.

It is generally found that the oscillations build up to such a value that the greatest positive potential of the grid, which will be represented by $E_{g\max}$, becomes practically equal to the lowest plate potential, $E_{p\min}$. It is not difficult to understand why this condition should represent a sort of limit. It is usually found that the dynamic characteristic, as shown at *B*, Fig. 56, tends to bend rapidly downward as $E_{g\max}$ becomes greater than $E_{p\min}$. When $E_{g\max}$ tends to become greater than $E_{p\min}$ the electron current to the grid rises suddenly (see cathode ray oscillogram, Fig. 57) with the result that the average

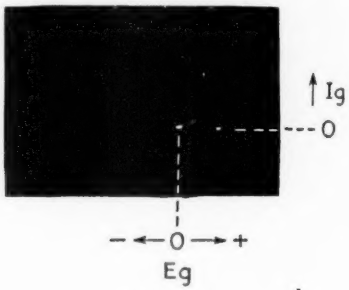


Fig. 57

current flowing through the grid leak increases very rapidly, which in turn results in a marked depression of the average grid potential.

The increased electron current to the grid as E_g tends to become greater than $E_{p\min}$ also represents a greater dissipation of energy upon the grid which is equivalent to an increase in the effective resistance *R*.

It is quite possible that any one of these three factors, in the absence of the others, would be sufficient to limit the oscillations; but as each springs into importance when the amplitude of the oscillation has reached about the same value, we shall not discuss the exact combination of the three which actually determines the amplitude.

Adopting the relation $E_{g\max} = E_{p\min}$ and making the additional assumption that the dynamic characteristic is straight (which is very nearly true) an expression for the amplitude of oscillation can readily

be obtained. Introducing certain approximations this expression may be written,

$$i_{p1}(R+R_o)=\frac{1}{\sqrt{2}}\,E_p, \tag{14}$$

in which terms of the order $\frac{1}{\mu}$ are neglected in comparison to unity. i_{p1} is the fundamental component of the space current, E_p is the mean plate voltage and the term R_o is largely determined by the tube, and while not generally equal to r_p, is apparently not much different from it. This equation indicates that the amplitude of oscillation is practically independent of the value of μ.

The remarkably simple relation given by Equation 14 has been tested for a wide variety of circuits and tubes and has been found to hold with a very fair degree of accuracy. It may safely be taken as indicating quite approximately what response may be expected from any tube and circuit when operated at a particular applied E_p. The equation is likely to be more closely followed the more carefully the adjustment of circuit for maximum efficiency has been made.

The condition $E_{g\max}=E_{p\min}$ should not be considered as invariable. Adjustments can readily be made for which $E_{g\max}$ will either be ap-

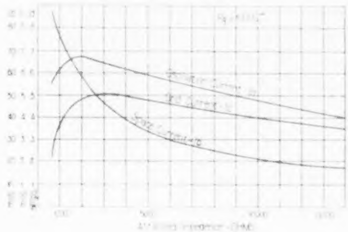


Fig. 58

preciably less or greater than $E_{p\min}$. It is generally found however that these adjustments do not give the most efficient operation.

Additional information as to the relation between the amplitude of oscillation and certain circuit constants are given in Figs. 58, 59, 60, 61. Fig. 59 shows that an oscillator tube may present a well-defined condition of temperature saturation. Figs. 60 and 61 show that the value of the feed-back voltage and grid leak resistance r_g may be varied within wide limits without affecting the output markedly.

49. *Efficiency*. The efficiency of an oscillator may be defined as the ratio between the energy of oscillatory current and the d.c. energy supplied to the plate circuit. This leaves out of account the energy

required to actuate the filament. The efficiency of oscillator circuits ranges all the way from a few per cent. to as high as 90% or better. The principal factors determining the efficiency are those which determine the amount of energy dissipated upon the plate of the tube. Inspection of the oscillogram, Fig. 55, shows that the sharper and

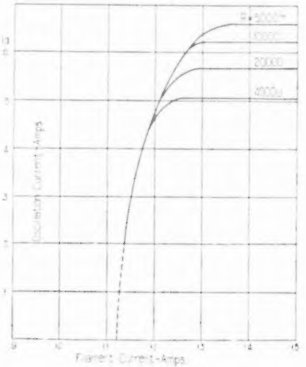


Fig. 59

narrower the plate current wave and the more nearly the plate voltage approaches zero, the higher will be the efficiency. In an extreme case such as the hypothetical one illustrated in Fig. 62 it is evident that the efficiency would be very large indeed. The μ of the tube

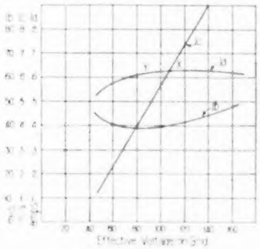


Fig. 60

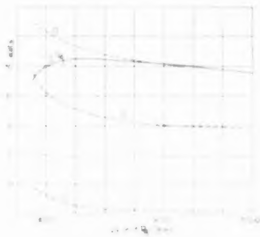


Fig. 61

largely determines the sharpness of the plate current wave and experience shows that for efficiencies of about 50% and better, μ should be at least ten; an increase above this value does not result in any very large improvement. It is also generally true that for the highest efficiencies the circuit constants should be so arranged that R is at

least four or five times as great as r_p, and it may advantageously be made 10 to 15 times as great. In these latter cases R_0 is relatively

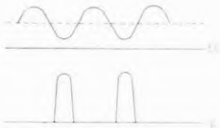


Fig. 62

negligible compared to R in Equation 14 with the result that i_{p1} can be quite accurately calculated although the exact value of R_0 may be unknown.

50. *Types of Oscillating Circuits.* There are many different types of oscillating circuits and as they do not lend themselves readily to classification, only a few of the more common types will be described.

One of the simplest oscillating circuits is that shown in Fig. 63 which is characterized by a tuned grid circuit inductively coupled

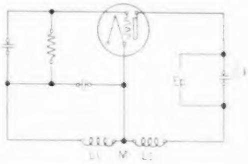


Fig. 63

to a coil in the output. This type is satisfactory for low plate voltages and small powers, but is not to be recommended where large amounts of power and high efficiencies are desired.

The condenser C_o is inserted to prevent short circuiting of the plate battery or generator and may be made so large as to have no effect on the frequency.

A circuit of similar properties is shown in Fig. 64, the output being tuned instead of the input.

51. *Colpitts and Hartley Circuits.* Two very similar types of circuits which have proved satisfactory for a wide range of frequencies, voltages and powers, and which yield very high efficiencies, are shown in Figs. 64 and 65, the former being known as the Colpitts and the latter as the Hartley circuit. In both circuits, as illustrated, the mean grid potential is secured by a grid leak. The blocking condenser C_g should be large enough to offer very little impedance to the flow of the alternating current which causes the variation of the grid potential,

and as seen in Fig. 61 the resistance of the leak can be varied within wide limits without an appreciable effect upon the performance of the circuit. The source of plate potential may be either battery or generator. The choke in the generator supply should be sufficiently large to prevent an appreciable amount of oscillating current flowing

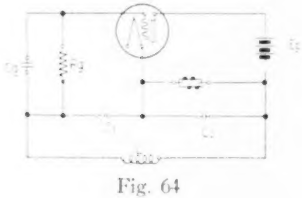


Fig. 64

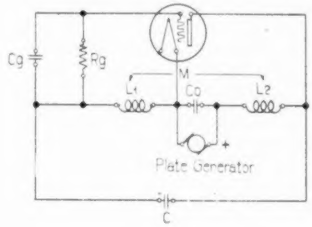


Fig. 65

through the generator. In order to avoid short circuiting the generator in the Hartley circuit, a large condenser C_o is interposed as shown in Fig. 65. For high frequencies it is customary to make this condenser so large that it has very little effect upon the frequency of oscillation.

The power of the oscillating current may be removed either by capacity or magnetic coupling, or it may be dissipated by resistance within the oscillating circuit. Magnetic coupling is very satisfactory for radio frequencies, but for the audio frequencies unless iron core transformers are used, capacity coupling usually proves more convenient. In regard to the use of iron core inductance and transformers, the reader should note the following Section.

52. *Frequency of Oscillation.* It is frequently found that the tube causes an appreciable deviation from the calculated frequency of oscillation. This is not so much because the tube influences the normal operation of the oscillating circuit as it is due to the fact that the impedances of inductance coils (particularly when iron cored) and of the condensers are liable to vary in value with the current amplitude. Hence any change in applied plate potential, filament current, or output resistance, because of its effect upon the amplitude of the oscillating current, will give rise to changes in frequency. In designing circuits to operate at a constant frequency regardless of slight changes in tube constants, plate potential, filament current, etc., the most important requirement is to provide inductances and condensers whose impedances are independent of the currents they carry, and whose resistance components are very small.

53. *Oscillator for A.C. Measurement Purposes.* For many a.c. measuring purposes an oscillator whose output is both free from harmonics and constant in frequency is desirable. Such a circuit is shown in Fig. 66, the design of which is radically different from the oscillator circuits already discussed. It possesses a tuned input *LC* and coupling is supplied by the resistance *R*. *R* is usually given a

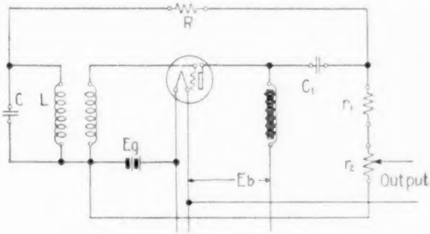


Fig. 66

value between 100,000 and 400,000 ohms. A negative grid battery fixes the average grid voltage, the emf. of this battery being about 8 to 10 volts. It is customary to make the resistances r_1 amd r_2 several thousands ohms apiece, r_1 being perhaps 5 times r_2. The condenser C_1 is merely a blocking condenser and should offer little impedance to the a.c.

It is not difficult to make the oscillator of Fig. 68 maintain a frequency that is constant within 3/10 of 1% and when the feed-back is not too large the harmonics in the output will comprise only 5% or even less of the total a.c. output.

54. *Range of Frequencies Obtainable with Vacuum Tube Oscillators.* Circuits have been constructed whose frequency is but a fraction

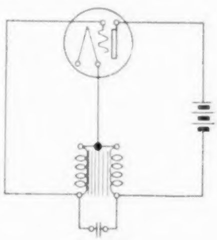


Fig. 67

of a cycle per second. The requirements of such a circuit are large inductance and capacity and very close coupling between input and output. A satisfactory circuit for low frequencies is that shown in Fig. 67; the two inductances taking the form of an iron core transformer.

At the other extreme, frequencies as high as 5×10^7 cycles per second can be obtained by means of tuned vacuum tube circuits of the Hartley or Colpitt type. At this point the coupling reactance of the tube becomes appreciable with that of the circuit.

Circuits capable of considerably higher frequencies have been described by Van der Pol,[37] Southworth,[38] Gutton and Touly,[39] and Holborn.[40] In all of these cases the oscillatory circuit is made up of distributed inductance and capacity connected to the tube in such a way as to utilize the capacity between the elements of the tube as a means of coupling.

The circuit shown in Fig. 68, when properly arranged, is as efficient as those used for lower frequencies and will give frequencies as high as 3×10^8 cycles per second. The oscillatory circuit is indicated by the heavy lines. It consists of a rectangle whose dimensions are appreciable with the wave length. Therefore, waves produced by variations in the electron emission through the grid are guided along the rectangle and are reflected at the ends. The reflected waves produce the proper voltage changes on the grid to sustain oscilla-

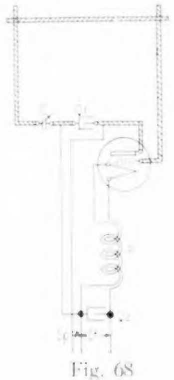


Fig. 68

tions. The ground imposed by the power leads places at least one point of the circuit at earth potential. If the condenser C is properly adjusted relative to the capacity between the grid and the plate the wave front can be made essentially perpendicular to the sides of the rectangle. It has been found that a large part of the power loss in the circuit is due to radiation. This circuit has been used as a basis

[37] B. van der Pol, *Phil. Mag.*, 38, July, 1919.
[38] G. C. Southworth, *Radio Rev.*, 1, Sept., 1920.
[39] Gutton and Touly, Comptes Rendus, 168, Feb. 3, 1919.
[40] F. Holborn, *Zs. fur Physik*, 6, p. 328.

of directive radio in which a metallic mirror was used to reflect the transmitted signals.

Very different means of producing high frequencies have been used by R. Whiddington,[41] Barkhausen and Kunz,[42] and by Gill and Morrell.[43] In some cases they employ tubes having considerable residual gas. The frequencies produced depend on the relative voltages applied to the grid and plate. Probably the best explanation of this phenomenon has been given by Gill and Morrell. Frequencies higher than 3×10^8 cycles per second have been reported.

The most accurate way of measuring these high frequencies is by observing the length of standing waves produced on a parallel wire system. The constancy of the vacuum tube generator, compared with spark oscillators, combined with the fact that the sharpness of resonance in a parallel wire circuit is comparable with that in ordinary radio circuits, makes it especially adaptable to measurement purposes. It may be used, for example, to measure small inductances and capacities or to determine the dielectric constant of liquids. Many of the corrections necessary when a damped source is used are eliminated.

55. *The Mechanically Coupled Oscillator.* In addition to the types of oscillators described above where the frequency is determined by

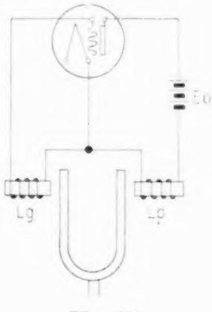


Fig. 68a

inductance and capacity, we may have oscillators in which the frequency is governed by a mechanical system such as a pendulum or a tuning fork.[44] An example is shown in Fig. 68a. The two coils

[41] R. Widdington, *Radio Rev.*, 1, Nov., 1919.

[42] Barkhausen and Kunz, *Phys. Zs.*, Jan, 1, 1920.

[43] Gill and Morrell, *Phil. Mag.*, 44, July, 1922.

[44] See Eckhardt, Karcher and Keiser, *J. O. S. A. & R. S. I.*, Vol. 6, p. 948, 1922; Eccles & Jordan, Phys. Soc. Proc. 31, Aug., 1919 and Phys. Soc. Proc. 32, Aug., 1920; Abraham & Bloch, J. d. Physique, Vol. 9, July, 1920.

L_p L_g are inserted in the plate and grid circuits of the tube. Variations in the plate current through the coil L_p impress forces on the tuning fork which result in its motion. This motion of the fork causes variations in the magnetic field through L_g and induces a varying voltage on the grid. With the proper coupling, sustained oscillations result having a period very nearly that of the tuning fork.

The electrically driven tuning fork described above constitutes a very satisfactory source of either sound or electromotive force. Horton, Ricker and Morrison [45] have made improvements which make it

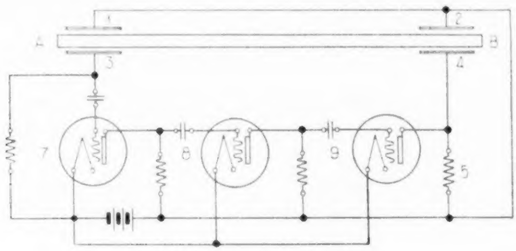


Fig. 68b

constant in amplitude and frequency to six parts in a million over very long periods of time.

An entirely different form of mechanical coupling has been used by Cady.[46] The circuit is shown in Fig. 68b. It makes use of the piezo-electric effect and mechanical vibrations of a crystal. Variations in the plate current in the tube 9 cause a voltage change across the resistance 5. This is communicated to a crystal AB such as quartz by means of the plates 2 and 4. A transverse electric field applied to such a crystal causes a change in its length. If this electric field be periodic, compression waves will travel along the crystal with a velocity depending on its density and elastic properties. These waves will, in turn, cause a varying electric field between plates 1 and 3 which may be communicated to the grid of the tube 7, amplified by 8, and finally transmitted to tube 9. This provides conditions for sustained oscillations having a frequency which is roughly inversely proportional to the length of the crystal.

Cady describes oscillators ranging in frequency from 3×10^4 to 10^6, and states that the frequency is constant to about one part in 10,000. The effect of temperature change is not great.

[45] *Journal of A. I. E. E.*, 1923.
[46] Cady, Proceedings of I. R. E., Vol. 10, No. 2, April, 1922.

X. Miscellaneous Applications of Thermionic Vacuum Tubes

56. *The Tube as a Voltmeter.* The three-element tube may be used for the measurement of either d.c or a.c. voltages. In the case of d.c. voltages it is customary to apply the unknown voltage to the plate, counter-balancing this voltage with a known negative potential applied to the grid. Given the μ of the tube, it is then possible to calculate with a fair degree of accuracy the plate potential. The usual procedure is to adjust the negative grid potential to such a point that the plate current just becomes zero. The tube when used in this manner becomes an electrostatic voltmeter, and it is evident that to give accurate readings the tube should have a well-defined cutoff (see Sec. 8, Fig. 10).

In a somewhat similar fashion a.c. peak voltages may readily be compared with known d.c. voltages. A typical circuit is shown in Fig. 69. In operation a fixed plate voltage is applied to the voltmeter

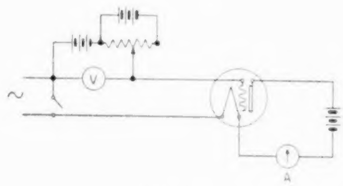


Fig. 69

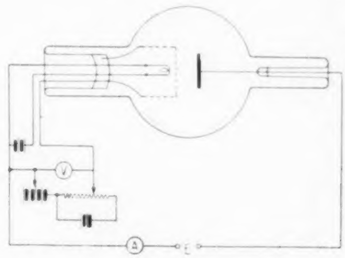


Fig. 70

tube and a steady negative d.c. voltage is applied to the grid which is just sufficient to reduce the plate current to zero. The a.c. voltage is then superimposed in the grid circuit with the result that current flows during the positive halves of the wave. If now the steady grid potential is made more negative until the plate current again just ceases to flow, it is apparent that this change in the steady potential just equals the peak value of the a.c. voltage.

For the measurement of very high voltages a special tube of the design shown in Fig. 70 will be found desirable, the grid being in the form of a screen which surrounds the filament. Such a tube may have a μ as high as 200.

A circuit similar to Fig. 69 may be so employed that the a.c. voltage to be measured causes a change in the space current meter reading, the negative grid potential being preferably so set that the conditions discussed in Sec. 16 are satisfied. The tube, due to its curved char-

acteristic, acts as a detector and as pointed out in Sec. 37, the change in space current is approximately proportional to the square of the a.c. input voltage. For accurate work the circuit requires calibration but the calibration will in general remain good over long periods of time. This method is particularly useful for small voltages.

57. *Power-Limiting Devices.* As pointed out in Sec. 4 the total emission from the filament at a given temperature is fairly sharply defined regardless of the plate voltage, so long as this exceeds the value required to give voltage saturation. The fact that the total emission is limited by the temperature may be used to control the

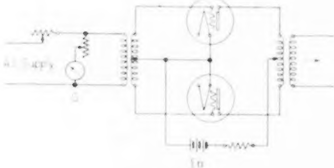


Fig. 71

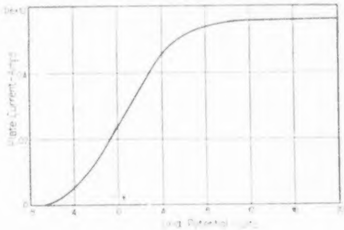


Fig. 72

maximum current in a circuit. As an illustration, Fig. 71 shows its application to an alternating current circuit, the performance of which is illustrated in Fig. 72. The introduction of such a device into an a.c. circuit will, of course, result in the generation of harmonics and may therefore, be objectionable.

There is almost no limit to the number of regulatory circuits which can be devised to employ the three-electrode tube.

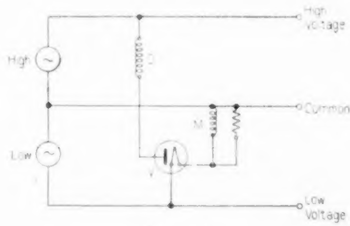


Fig. 73

58. *Voltage and Current Regulation of Generators.* The two-electrode tube with tungsten filament has been used to great advantage as a voltage regulator for a special airplane generator designed to deliver both 28 volts and 300 volts. The circuit arrangement is illustrated

in Fig. 73 in which M represents the main field winding, and D the differential winding which opposes M. The high voltage given by the generator when applied to the plate of the valve is sufficient to produce a condition of voltage saturation. As the speed of the generator increases, the current through the main field winding and

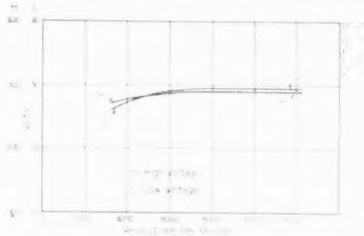


Fig. 74

the filament increases, thereby giving rise to greater emission from the filament and a larger current through the differential winding. It was found possible to so design the valve as to yield the very close regulation illustrated in Fig. 74.[47]

The three-electrode tube can also be used as a voltage regulator for a generator as shown in Fig. 75. It is apparent that an increase

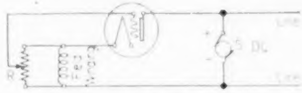


Fig. 75

in the voltage across the line tends to increase the current through the tube and resistance R. This in turn lowers the grid potential and tends to prevent an increase in current through the field winding.

The circuit shown in Fig. 76 illustrates an arrangement for main-

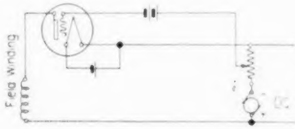


Fig. 76

Fig 77

taining a constant current from a generator. The operation of the device is apparent.

Fig. 77 shows another arrangement for maintaining a constant generator voltage. In this circuit an increase in voltage tends to

[47] Radio Telephony by Craft & Colpitts, Trans. A. I. E. E., Vol. 38, p. 330, 1919

make the grid less negative, thereby reducing the resistance shunted across the field winding.

A somewhat similar arrangement can readily be applied to regulate the voltage delivered by a battery. Fig. 78 illustrates such a circuit. An increase in E_1 raises the grid potential, thereby increasing the current through the tube and the resistance r_3. By a choice of regulating tube and resistances such that

$$r_3 = \frac{r_1 + r_2}{r_1} \frac{dE_g}{dI_p},$$

it may readily be shown that the voltage E_2 remains constant. Since the regulation effected by this circuit is independent of frequency it

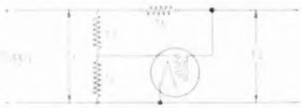


Fig. 78

may also be applied to a generator supply for elimination of commutator noise as well as voltage fluctuations due to changes in speed.

59. *The Ionization Manometer.* When gas is present in a three-electrode tube in quantities not sufficient to seriously affect the activity of the filament, and the plate voltage exceeds a value sufficient to produce ionization by collision, it has been found that the number of ions produced is proportional both to the pressure of the gas and to the electron current passing through the gas to the anode.[48] If now a small negative potential be applied to the grid, a certain fraction of the positive ions will be drawn to it and their number can be accurately measured by the current flowing in the grid circuit. The best arrangement is to apply the positive potential, not to the plate in the usual fashion, but to the grid, and apply the negative potential to the plate making it the collector of the positive ions. Dimensions of a satisfactory tube are given in Fig. 79. The values $E_g = 110$ volts and $E_p = -2$ volts have been found to give very satisfactory results, the electron current being .02 ampere, and K being equal to 0.10 for nitrogen and having approximately this value for air. The gauge equation may be put in the form,

$$P = K \frac{I+}{I-},$$

in which P is the pressure, K a constant depending upon the design

[48] O. E. Buckley, Procd. Nat. Acad., Vol. 2, p. 683, 1916.

of the tube, $I+$ the positive ion current, and $I-$ the electron current.

As it is necessary to know the value of $I-$, and since the emission from the filament is liable to vary somewhat with the kind of gas and its pressure, it will be found advantageous, if many readings are to be made, to place in the $I-$ circuit, the coils of a relay which is adjusted to close at a definite value of $I-$, and which, when closed, cuts in a shunt around the filament which will reduce its heating current.

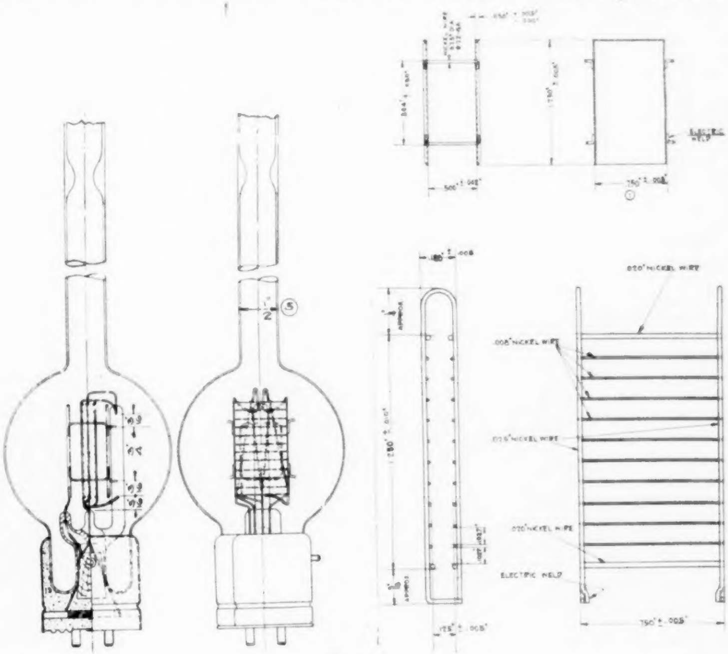


Fig. 79

Automatic regulators of this type have been used with complete success. Experiment shows that the value of K remains constant for pressures as high as 1.5×10^{-3} mm. of Hg. and the lower limit is determined very largely by the sensitivity of the current reading instruments. It follows that an ionization gauge can be calibrated by comparison with a McLeod gauge and, in use, extrapolated to low pressures.[49]

60. *Heterodyne Method of Generating Currents of Very Low Frequencies.* By impressing upon the grid of a detector tube two frequencies which differ by a very small amount (e.g., 99 and 100 cycles),

[49] Dushman, *Phys. Rev.*, Oct., 1920, p. 854.

it is possible to obtain from the output of the detector the difference frequency of one cycle per second. This low frequency may be readily separated by means of a filter. It is apparent, however, that to maintain this difference frequency constant requires that the input frequencies be held within a very narrow range of variation.

61. *Thermionic Valve as a High Tension Switch.* If the plate circuit of a valve is inserted in a high tension circuit, the flow of current in the circuit may readily be stopped by cutting off the filament heating supply, thus making unnecessary the breaking of any contacts in the high tension circuit. In case the transmission of current in both directions is necessary, two valves may be used.

62. *Devices Employing Secondary Emission.* As pointed out in Sec. 9, the grid current in a three-element tube shows a negative resistance characteristic for a certain range of voltage, and various uses of this fact have been pointed out.[50]

63. *Electron Tube Oscillograph.* A special type of thermionic tube designed for oscillographic uses is of great importance as a laboratory instrument. These tubes, using the hot filament as a source of electrons, have certain marked advantages over the Braun tube with its gaseous discharge.[51] One of the very successful thermionic oscillographs has the following properties: anode potential 300–400 volts, sharp focus of electron beam, sensitivity of 1 mm. per volt between deflection plates and 1 mm. per ampere-turn when using magnetic deflection. Photographic recording is possible with relatively short exposures by using suitable fluorescent material.

[49] Dushman, *Phys. Rev.*, Oct., 1920, p. 854.
[50] See footnote 21.
[51] See J. B. Johnson, *J. of Opt. Soc. of Amer.*, Sept., 1922, or *Bell System Technical Journal*, Nov., 1922.

Some Contemporary Advances in Physics

By K. K. DARROW

NOTE: Dr. Darrow, the author of the following article, has made it a practice to prepare abstracts and reviews of such recent researches in physics as appear to him to be of special interest. The results of Dr. Darrow's work have been available to the staffs of the Bell System laboratories for some time and having been very well regarded, it is thought that such a review, published from time to time in the TECHNICAL JOURNAL, might be welcomed by its readers.

The review cannot, of course, cover all the published results of physical research. The author chooses those articles which appear significant to him or instructive to his readers, without attempting to pass judgment on the scientific importance of the different papers published. It is not intended that the review shall always assume the same form; at one time it may cover many articles, at another be devoted to only a few, and it may occasionally treat of but a single piece of work.—*Editor.*

SOME years ago C. T. R. Wilson of Cambridge University developed a beautiful method for making the paths of moving charged atoms and electrons individually visible. The charged particle flies through a gas such as air, mixed with water-vapor; it ionizes many of the molecules near which it passes; the gas is suddenly cooled by expansion and the water-vapor is precipitated upon the ionized molecules, forming a trail of droplets which visibly mark out the path of the ionizing electron or atom. Truly spectacular photographs of such trails, thick straight ones of fast-moving atoms and thin curly ones of electrons, are frequently published in textbooks and in popular articles.

The method is now proving very powerful in the study of collisions and close encounters of electrons with atoms and of atoms with atoms. Rutherford having found by another method that the nuclei of atoms are occasionally broken up by unusually direct blows from fast-moving helium nuclei (alpha-particles), the prospect of actually photographing such an important event becomes alluring. However, it is a very rare event; for W. D. Harkins and R. W. Ryan of the University of Chicago photographed eighty thousand alpha-particle trails in air, and only three of the particles struck molecules so squarely as to be deflected through more than a right angle; and of these only one showed indications of having broken the nucleus it struck. This particular collision is shown in Fig. 1 (two photographs of the same encounter taken from different directions at the same moment). In addition to the tracks of the alpha-particle up to and away from the scene of the encounter, there are two more tracks diverging from it, which are probably the tracks of two fragments of the struck nucleus. Other interpretations, such as two distinct impacts very near together or a stray radioactive atom

happening to disintegrate just as the alpha-particle passed by, are admissible but highly improbable. Another such collision in argon is shown in Fig. 2; this too was the only encounter with four diverging

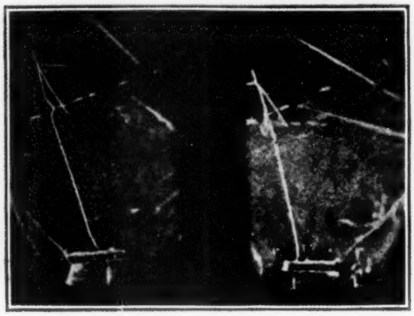


Fig. 1

tracks observed in many thousand photographs with the same gas.[1] A collision in air, in which the struck nucleus was not broken, but knocked to one side while the alpha-particle rebounded in the manner demanded by the principle of conservation of momentum, is shown

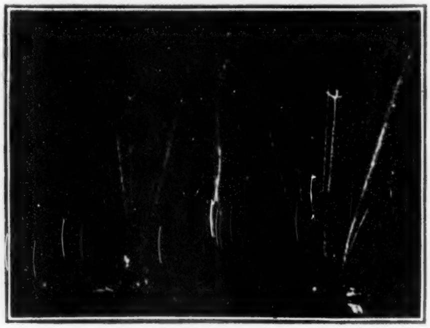


Fig. 2

in Fig. 3. These results show how small the atom-nuclei must be, compared to the extension of their electron-systems; for the 80,000 alpha-particles observed in air had traversed the electron-systems of about ten billion molecules altogether.

[1] As Rutherford's experiments indicate that argon atoms are especially stable against disintegration, this may be a case of two consecutive collisions with adjacent atoms.

Fig. 4 shows curious collisions of alpha-particles passing through helium gas, photographed by D. Bose and S. Ghosh of Calcutta. In each of the two left-hand trails the alpha-particle has apparently

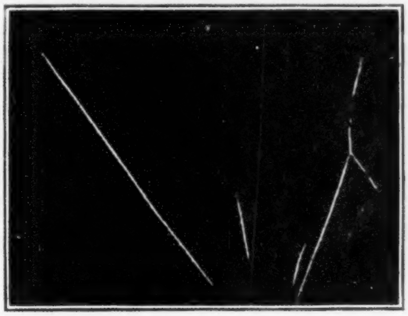


Fig. 3

knocked the nucleus and the two electrons of the atom in three different directions.[2] The alpha-particle of the right-hand trail (*iiib* is a magnification of *iiia*) seems to have produced quite an explosion; this may be the disruption of a nucleus belonging to a stray molecule

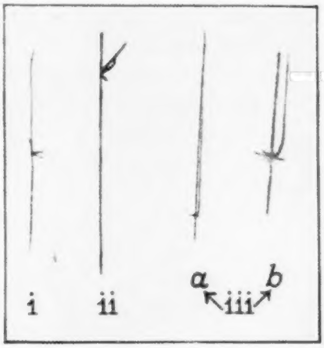


Fig. 4

of nitrogen, but one would not expect the original particle to go on as if unaffected.

[2] In hydrogen they found no cases of two electrons being driven off by the same impact. This agrees with Millikan's conclusion that double ionization is much more frequent with helium atoms than with molecules of any other kind, if indeed, it is not a characteristic of helium alone.

Much work is now being devoted to the spectra of ionized atoms. One instance, that of ionized potassium, will suffice to illustrate the problem. The potassium atom differs from the argon atom in three respects; the weight of its nucleus is slightly different, which is probably inessential; the charge on its nucleus is 19/18 as great; and it has a nineteenth electron outside of the three closed electron-shells, comprising eighteen electrons, which by themselves constitute the whole electron-system of the argon atom. When this outermost electron is removed, we have a system which probably differs from the argon atom in only one essential respect—that the central nucleus has a somewhat larger attractive power and hence the three electron-shells are somewhat more drawn inward. The spectra of ionized potassium and of argon should therefore be very nearly alike. This has been tested by Zeeman and Dik at Amsterdam; the result is very satisfactory, and the difference between the simple and clearly-arranged spectrum of potassium on the one hand, and the rich and intricate spectra of ionized potassium and argon on the other hand, is very striking. At Bonn, the spectrum of ionized rubidium is being compared with that of krypton for the same purpose.

The most extensive results, however, have been obtained by Fowler with silicon. For some reason or other, silicon is a particularly easy element from which to obtain spectra not only of the neutral and the ionized atom, but also of the twice-ionized and thrice-ionized atom—four distinct spectra, one from neutral silicon, the next from an atom resembling aluminium, the next from an atom resembling magnesium, and the last from an atom resembling sodium. These four spectra can be observed in the stars and in the laboratory, some of the important lines from thrice ionized atoms having been photographed by Millikan in the extreme ultra-violet. In their general type, they resemble the spectra of the neutral atoms corresponding in structure to the atoms which emit them.

Data have also been made available for doubly-ionized magnesium (by Paschen) singly-ionized magnesium, and for neutral sodium—three atoms in which the nuclear charge is respectively 13e, 12e, and 11e, while in each of them the nucleus is surrounded by ten electrons and there is an eleventh one much further out. This eleventh electron being responsible for the spectrum and being relatively exempt from perturbations due to the other ten, the spectra of these three atoms are of the simplest and clearest type. The series-lines which in the spectrum of neutral sodium are in the inaccessible infra-red are moved up, in the spectrum of doubly-ionized aluminium, into the visible region. Further study of spectra related to each other in this

manner, and differing by virtue of slight intelligible differences in the atoms which emit them, may be expected to help greatly in making clear the major features of atomic structure.

Two phenomena, first accurately examined by A. H. Compton, afford a striking illustration of the way in which classical electromagnetic theory and quantum theory are alternately successful in explaining the qualities of radiation. On the one hand, Compton has been the first to apply accurate wave-length measurements to scattered X-rays, and finds that they are a mixture of two kinds of X-rays—one having exactly the same wave-length as the primary X-rays, the other a wave-length slightly greater and varying with the angle between the primary and the secondary rays. According to the classical theory, scattered X-rays are simply radiation sent out in all directions by electrons inside the atoms of the scattering substances, vibrating under the influence of the primary X-rays, and hence vibrating necessarily with the same frequency as the primary X-rays. This could account for one of the components of the scattered X-rays, but not the other. The other can be accounted for by assuming that the primary X-ray quanta of frequency n are perfectly elastic spheres which travel with the velocity of light, have momentum hn/c and energy hn, and collide with the atoms just as one elastic sphere collides with another (that is, under conditions of conservation of translatory kinetic energy and of momentum); they depart from the collision with less energy and less momentum than they initially had, and consequently with a diminished frequency. But this does not explain the first-mentioned component, leaving the two theories balanced. On the other hand, in the *Philosophical Magazine* paper, Compton describes the total reflection of X-rays by glass, silver and lacquer—a phenomenon of exactly the type which the classical theory explains far more easily and naturally than the quantum-theory.

In glass and lacquer, the highest natural frequency of any of the electrons in any of the constituent atoms—to speak the language of the classical theory—is far below the frequency of available X-rays; we are, in optical terminology, on the high-frequency or anomalous-dispersion side of the highest-frequency absorption-band; the well-known dispersion formula reduces to a single term,

$$\mu = 1 - Ne^2/2\pi mn^2$$

where N is the total number of electrons able to vibrate in unison with the X-rays, and μ is the index of refraction of the X-rays of

frequency n; e and m have their usual meanings. The index of refraction is less than unity, the X-rays travel faster in glass or in lacquer than in air or in vacuo, and are totally reflected from a glass surface if incident at a sufficiently small angle with the surface. The agreement between experiment and theory is, quantitatively as well as qualitatively, very good. It is equally good for silver, allowance being made for the fact that the frequency of the X-rays used lay between the two absorption-bands of silver. It seems conceivable that this might be refined into a method for determining the numbers of electrons in different orbits of the atom.

The atoms of the inert or "rare" gases argon, krypton, and xenon are almost completely transparent to slow electrons—electrons moving with a speed of one or two equivalent volts. In more exact language, the radius of the effective cross-section of one of these atoms relatively to slow electrons is much smaller than its radius relatively to faster electrons or to other atoms. This almost incredible statement, having been tested by several different experimenters and by at least two entirely distinct methods, now appears to stand beyond doubt. This radius of the effective cross-section of the atom, relatively to an electron, is (by definition) the least distance at which the electron can pass by the centre of the atom without being intercepted or deflected; the radius of the atom relatively to another of the same kind is, naturally, half the least distance at which the centres of the two atoms can pass each other without affecting one another's paths. The concept is not perfectly exact, depending as it does on what we choose to take as the least perceptible alteration of the path of a particle; nevertheless, it is practicable and useful. Years ago the radius relative to other atoms was determined (from the viscosity of the gas). There is no binding reason why it should be identical with the radius relative to electrons, but the first measurements of this latter quantity on such gases as hydrogen, nitrogen, and helium yielded fairly good agreements between the two. Recent measurements on argon disclosed a surprising difference.

The method consists essentially in measuring the fraction of a beam of electrons, projected against a layer of gas, which pass through the layer undeflected. (Another and entirely different method used by Townsend resulted in a valuable confirmation of the result.) If there are N atoms under unit area of the surface of the layer (looking through it in the direction from which the electrons come) and N is not so large that many of the atoms are partly shielded, in the perspective, by others, the fraction of the electrons which go through undeviated is $(1-N\pi r^2)$; r being the radius just defined. The most

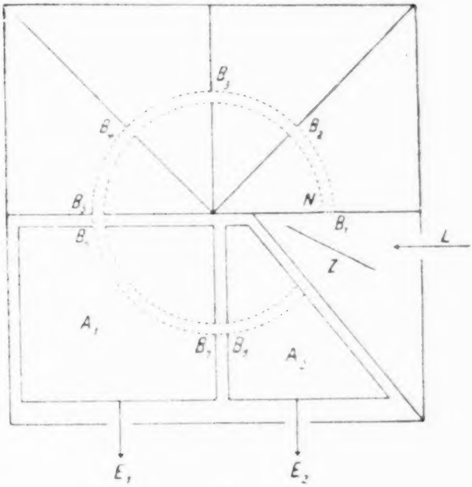

B3
B4
B2
B5
N
B6
B1
L
Z
A1
B7
B8
A2
E1
E2

Fig. 5

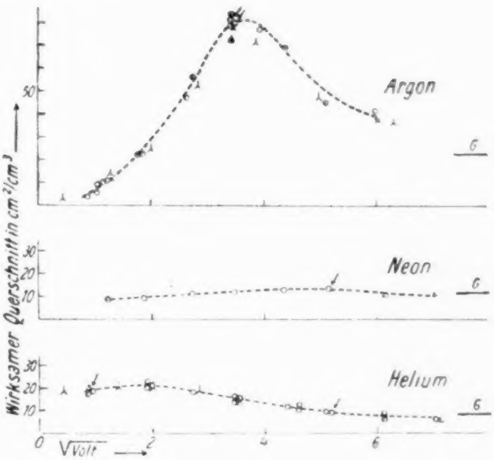

Argon
Neon
Helium
Wirksamer Querschnitt in cm²/cm³
√Volt
50
30
20
10
0
2
4
6
G

Fig. 6

delicate arrangement is that of Ramsauer (Fig. 5); the electrons enter at B_1, and are steered by a magnetic field along a semicircular path through the slits B_2-B_8; electrons deviated even through a very slight angle go against the partitions and are not received by the electrometers connected at E_1 or E_2. Measurements with the two

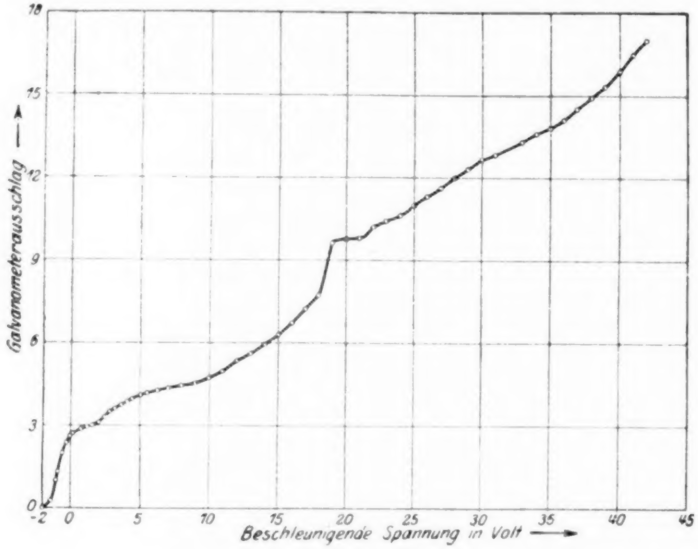


Fig. 7

electrometers, at two pressures of the gas, yield the data required. The values of r^2 thus determined for argon, neon and helium are plotted against the speed of the electrons in Fig. 6. The ordinates of the short straight horizontal lines on the right represent the squares of the radii relatively to other atoms.

The effect shows itself, however, very distinctly in a much simpler and more common device, a cylindrical three-element tube of audion type with the grid very much closer to the filament than is the plate; the plate is maintained at a potential a fraction of a volt higher than that of the grid. Figs. 7 and 8, from a recent article by Minkowski and Sponer, exhibit curves of plate-current versus grid-voltage in helium, which does not show the effect in question, and argon, which does.[3] In helium the current rises steadily as the increasing voltage

[3] The displacement of the curves by about −2 volts along the axis of voltages is probably due in part to drop of potential along the filament, in part to neglected contact-potential-differences.

gradually overcomes the space-charge repulsion, augmented in the gas by the reflection of electrons, for the reflected electrons stay longer in the space between filament and plate than they would if they went straight through. In argon the curve rises at first more swiftly, almost or quite as steeply as in vacuo, for the atoms are almost transparent to the electrons when they are slow; but as their speed is increased and the effective radius of the atom rises, the current sharply declines again. Further on, near 11 volts, there is

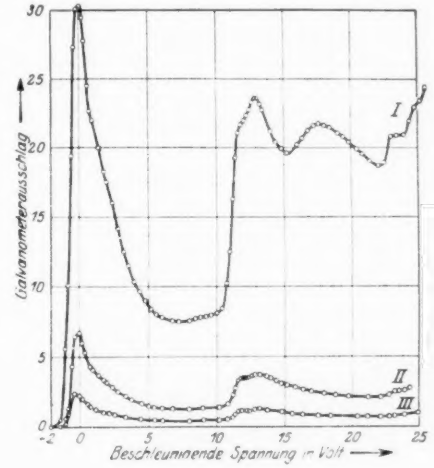


Fig. 8

another peak; near this voltage, the electrons which collide with atoms lose almost all their energy (threshold-speed for inelastic impact at 11.3 equivalent volts) at the first collision, and pass through the rest of the gas-filled region without obstacle. A second peak near 16 volts is ascribed to a second critical speed for inelastic impacts. Krypton and xenon give toothed curves of the same general type. Neon and mercury vapor, however, behave like helium, the curves rising steadily or at most showing slight kinks and inflections which may be indications of a slight effect of the same sort.

The reason for this remarkable effect is still obscure. It may be possible to devise an atom-model adequate to explain it without forfeiting spherical symmetry, which it is desirable to retain if possible, for among all atoms these of the heavy inert gases would be

expected to display the most complete symmetry and smoothness. F. Hund tried to devise an atom such that any one-volt electron passing within a distance r of the centre would be deflected through exactly 360° before coming out; he attained a formal solution of the problem, but the model involved a continuous distribution of negative charge from the nucleus outward to the distance r, which is quite incompatible with all our other knowledge of atomic structure. H. A. Wilson of Rice Institute tried out a well-known and popular model, consisting of a nucleus surrounded by a spherical surface of radius r, over which negative charge equal in amount to the positive charge on the nucleus is uniformly spread. For very slow electrons, the average angle of deflection is 90°; it increases with speed, becoming 180° at a certain critical value v_0 at which every electron is turned back into the direction whence it came; beyond v_0 it decreases indefinitely with increasing speed. At v_0 the oncoming electrons are more radically deflected by the atoms, so to speak, than at any greater or lesser speed; below v_0 the variation of mean deflection with speed is in the proper sense to agree with experiment, but not by any means of a sufficiently great order-of-magnitude. The theory, however, seems to explain the mild variations encountered in such gases as hydrogen and nitrogen, and the explanation of the more striking ones may lie in the same direction.[4]

Important contributions have lately been made to our knowledge of self-sustaining discharges, such as the glow and the arc, which maintain themselves as long as the proper voltage is applied at the electrodes, without requiring the assistance of a separate source of ions such as a hot filament or an outside ionizing agency such as X-rays. The field has perhaps been somewhat neglected, because it is easier to obtain simple clear results with electrons and ions admitted into a very rarefied gas after being generated elsewhere. In a self-sustaining discharge, there is usually a sudden steep potential-drop just in front of the anode, and another just in front of the cathode—the so-called anode-fall and cathode-fall; in the region between, the potential varies gradually. The anode always tends to become very hot, and Gunther-Schulze at Berlin has lately measured the rate at which heat is generated at the anode of a mercury arc; he finds that it agrees wonderfully well with the rate calculated from the assumption that practically the entire current is carried by negative ions

[4] Any competent theory must explain the results obtained by different methods, notably the fact that the value of r measured in an apparatus like Ramsauer's agrees with the value measured in an apparatus in which electrons deflected through considerable angle should yet reach the collector, and so be counted as though they had not been deflected at all.

(probably electrons) which dash against the anode with the entire kinetic energy acquired during unobstructed passage through the anode-fall.[5]

The heat generated at the cathode must arise in the converse way, from the kinetic energy of positive ions pulled violently against the cathode surface. K. T. Compton of Princeton, has made elaborate calculations for the arc in air with a carbon cathode and the arc in hydrogen with a tungsten cathode, and comparing the results with the experimental evidence, concludes that a few per cent of the current at the cathode is carried by positive ions, the remainder by electrons moving away from the cathode. Compton then attacked the same problem in an entirely different manner; he assumed that the region near the cathode, in which the cathode-fall occurs, is a region in which positive ions are moving gradually towards the cathode, accelerated by the field, and retarded by their collisions with neutral molecules and by their mutual space-charge repulsion. The problem is formally similar to that of determining the current-voltage relation in a thermionic vacuum-tube, and the solution is a relation between cathode-fall, current, and width of the region in which the cathode-fall occurs. The first quantity is known; the third is assumed to be the mean distance which an electron travels from the cathode before striking a molecule; the second quantity, the current of positive ions into the cathode, comes out to be a few per cent of the observed total current. These two methods thus support one another in indicating that in the arc-discharge some 90–98% of the current near the cathode is carried by electrons, and the small remainder by positive ions. In the glow-discharge, according to experiments by Gunther-Schulze, the rate at which heat is generated at the cathode is 25% to 75% of what it would be if all the current were carried by positive ions, falling against the cathode with the entire energy derived in passing through the cathode-fall. Expecting that a much larger fraction of the energy of the positive ions would be dissipated in collisions with neutral gas molecules, he concludes that the region of the cathode-fall must be a region in which the gas is abnormally rarefied because abnormally hot; the hotness in turn being due to the collisions between ions and molecules.

In the central region of the arc, the potential-gradient is uniform and consequently the positive and negative charges per unit volume

[5] It is obviously necessary to be very cautious in making deductions of this kind, for the entire energy iV (i representing the current and V the anode-fall or cathode-fall, as the case may be) is dissipated as heat in the region of the anode-fall or cathode-fall; and if this region is very narrow it is hard to distinguish between heat generated within it and heat generated at the anode or cathode surface.

must exactly balance one another. In this region Compton suggests that the gas is in the state of thermal ionization defined and described by Saha, in which at all times a certain constant percentage of the atoms, depending only on their ionizing-potential and on the temperature, is ionized. If the temperature of the central region of the carbon arc is about 4000°, and the ionizing-potential of the gas about 8 volts, the proportion of ionized molecules will be about right.

According to one of the newer and stranger developments of the quantum-theory, an atom possessing magnetic moment and submerged in a magnetic field is not at liberty to orient itself in any direction whatever, not even momentarily; it may set itself only at certain specified inclinations, such that the cosine of the angle between the direction of its magnetic axis and the direction of the field will have one of certain specified values. Imagine for example, an atom consisting of a single electron revolving in a one-quantum orbit (the smallest possible orbit) about a centre which itself is not magnetic; such a centre might be a simple nucleus, or a nucleus surrounded by a number of electrons moving in orbits so inclined to each other that their magnetic moments cancel one another out. The magnetic moment of such an atom is $eh/4\pi m$ (e the charge and m the mass of the electron); its magnetic axis is perpendicular to the plane of the orbit of the electron. According to the theory, the magnetic axis must point exactly with or exactly against the magnetic field; the cosine of the angle must be $+1$ or -1. This was verified last year by Gerlach, who projected a ray of silver atoms (shooting off from a hot rapidly-evaporating silver filament through a small hole) across a magnetic field with an extremely steep field-gradient. The ray divided itself into two, one consisting of atoms with their north magnetic poles pointing directly up the field, the other of atoms turned through 180° relatively to the first set; there was quantitative agreement with the theory. If the outside electron moves in a two-quantum orbit, the magnetic moment of the atom is $2\,eh/4\pi m$, and the cosine of the angle may take the values ± 1 and the values $\pm\frac{1}{2}$; if in a n-quantum orbit, the moment is $neh/4\pi m$ and the permissible values for the cosine are $\pm 1/n,\ \pm 2/n, \ldots\ldots \pm n/n$.[6]

The theory also accounts for the normal Zeeman effect. It remains to be settled whether the magnetic moments of actual paramagnetic substances can be calculated from it. According to the accepted

[6] The condition governing the angle is, that the integrals of (a) the angular momentum of the electron in its orbit, and (b) the projection of the angular momentum on the plane normal to the field, taken around a complete cycle of the orbital motion, must both separately be integer-multiples of the quantum-constant h. The latter integer-multiple cannot be zero, according to Gerlach's experiment and Sommerfeld's theory.

belief, the atoms of a paramagnetic substance all have a given constant magnetic moment, but are oriented in every possible direction so that the resultant magnetic moment of any piece of the substance is zero. If all the atoms could be made to point in the same direction by a powerful magnetic field, the total moment of the piece would be equal to the number of atoms in it multiplied by the moment of each atom, which could then be determined. No attainable magnetic field is strong enough to do this; the persistent effort of the field to twist the atoms into parallelism is almost completely counterbalanced by the thermal agitation. The total moment of the piece when all the atoms are parallel, and therefore the moment of each atom, have therefore to be calculated from the trend of the magnetization-versus-field strength curve in its attainable portion. In making this calculation it has heretofore been assumed that all orientations of the atoms are possible. Replacing this assumption by the contrasting one explained in the foregoing, we find the method of calculation altered;[7] the data heretofore assembled remain valid, but the values of magnetic moment computed from them are replaced by an entirely new set.

The old set of values of magnetic moment, calculated for a number of solid and gaseous substances and of ionized liquids, by Weiss and others, were said to be integer multiples of a fundamental constant, the "Weiss magneton." No one had succeeded in calculating the observed value of this constant from any atomic theory, and it is not compatible with the picture of the atom given above. The new set of values, according to Gerlach and to Pauli, who have worked over the published experimental material, is compatible with the atom-model. The values for solid platinum and palladium; for nickel in its high-temperature non-ferromagnetic "beta" form; and for nitric oxide gas, agree with the simplest model—the electron in a one-quantum orbit revolving around a non-magnetic centre. The value for gaseous oxygen agrees with the model having an electron in a two-quantum orbit; gamma-iron with the three-quantum, Mn_2O with the 4-quantum and MnO with the 5-quantum model. Various ions in solution from Cabrera's data also give values in accordance with the theory. It is implied that these cover all the reliable ob-

[7] In the latter case it is assumed that the number of atoms oriented with their axes in one permissible direction D_1 stands to the number oriented in another permissible direction D_2 in a ratio given by $exp\ (W/kT)$ where T is the temperature, k is Boltzmann's constant, and W is the work required to twist an atom from direction D_1 to direction D_2 against the magnetic field. In the former case all directions are regarded as permissible, and in the assumption just stated, "number of atoms oriented in direction D" is replaced by "density in solid angle of atoms oriented in direction D," a fundamental change.

servations on paramagnetic substances, except for two ions which yield values not reducible to agreement with the new theory.[8] The new method of calculating magnetic moments thus leads to values which confirm the contemporary atom-model. It would not be desirable to dismiss the old method and the old theory too hastily, considering that they lead to values which are claimed to be integer multiples of an apparently fundamental constant; but this constant has proved so intractable to theory that it would be gratifying to be able to discard it.

The arrangement of atoms in two samples of Heusler alloys was investigated with the X-ray method by J. F. T. Young at Toronto. These alloys are mixtures of the metals, copper, manganese, and aluminium in certain proportions; they are strongly ferromagnetic while the component metals are not ferromagnetic at all. Of the two samples, one had a much higher permeability than the other; the atoms of the former sample were arranged in a body-centered cubic lattice, with no trace of the characteristic lattices of the component metals. The atoms of the latter sample were arranged in a face-centred-cubic lattice. Thus these alloys furnish an additional instance of the frequent, though not by any means universal, correlation between body-centred-cubic lattice and strong ferromagnetism. L. W. McKeehan of the Western Electric used the same method to investigate palladium containing great quantities of occluded hydrogen. The space-lattice of the hydrogen-free metal was distended by a certain fixed percentage by saturating it with hydrogen; and it appeared that when the palladium contained a lesser quantity of hydrogen than the maximum or saturation amount, some parts of it were quite saturated and others contained no hydrogen at all, instead of the whole lattice being equally enlarged; it is probable that the individual crystals of the metal are saturated one by one as the hydrogen creeps in.

[8] The value for beta-iron as quoted by Gerlach does not agree with the theory. As for the values assigned by Weiss to the three ferromagnetic metals iron, nickel and cobalt, obtained from direct measurements of the saturation-intensity at the temperature of boiling hydrogen, the first two do not agree with the theory, the last agrees very well (assuming the electron to be in a one-quantum orbit). Of course, it is likely enough that the theory should not be applied to ferromagnetics. It seems fitting to quote a remark of Andrade about theories of magnetism in general " . . . the substances selected for verification of theories are of a very limited class, called of normal behavior rather because they agree with the theories than because they represent a numerical majority."

REFERENCES

D. M. Bose and S. K. Ghosh: *Nature* 111, pp. 463–464; 1923.

A. H. Compton: *Physical Review*, 2d ser., 21, pp. 483–502 and 715; 1923. (X-ray scattering.)

A. H. Compton: *Philosophical Magazine*, 43, pp. 1121–1129; 1923. (Total reflection of X-rays.)
K. T. Compton: *Physical Review*, 2d ser., 21, pp. 266–291; 1923.
W. Gerlach: *ZS. fur Physik*, 9, pp. 349–355; 1922. (Magnetic moment of silver atoms.)
W. Gerlach: *Physikalische ZS.*, 24, pp. 275–277; 1923.
H. Gunther-Schulze: *ZS. fur Physik*, 15, pp. 8–23; 1923. (Cathode-fall.)
H. Gunther-Schulze: *ZS. für Physik*, 13, pp. 378–391; 1923. (Anode-fall.)
W. D. Harkins and W. H. Ryan: *Journal Am. Chem. Soc.*, 45, pp. 2095–2107; 1923.
F. Hund: *ZS. für Physik*, 14, pp. 241–263; 1923.
L. W. McKeehan: *Physical Review*, 2d ser., 21, pp. 334–342; 1923.
R. Minkowski and H. Sponer: *ZS. fur Physik*, 15, pp. 399–409; 1923.
F. Paschen: *Annalen der Physik*, 71, pp. 142–161; 1923.
W. Pauli: *Physikalische ZS.*, 21, pp. 615–617; 1920.
C. Ramsauer: *Annalen der Physik*, 66, pp. 546–558; 1921.
H. A. Wilson: *Proc. Royal Society*, A103, pp. 53–57; 1923.
J. F. T. Young: *Philosophical Magazine*, 46, pp. 291–305; 1923.

Transatlantic Radio Telephony[1]

By H. D. ARNOLD and LLOYD ESPENSCHIED

SYNOPSIS: The first transmission of the human voice across the Atlantic was accomplished by means of radio in 1915. Since that time substantial progress has been made in the art of radio telephony and in January of this year another important step was taken in the accomplishment of trans-oceanic voice communication. At a prearranged time telephonic messages were received in London from New York clearly and with uniform intensity over a period of about two hours.

These talking tests were part of a series of experiments on transatlantic telephony which are now under way, the results of which to date are reported in this paper.

A new method of transmission, radiating only a single side-band, is being employed for the first time. As compared with the ordinary method of transmission, this system possesses the following important advantages:

The effectiveness of transmission is greatly increased because all of the energy radiated is effective in conveying the message; whereas in the ordinary method, most of the energy is not thus effective.

The stability of transmission is improved.

The frequency band required for transmission is reduced, thus conserving wave length space in the ether and also simplifying the transmitting antenna problem.

An important element of the high-power transmitter is the water-cooled tubes, by means of which the power of the transmitted currents is amplified to the order of 100 kilowatts or more. The direct-current power for these tubes is supplied from a 60-cycle, a-c. source through water-cooled rectifier tubes.

A highly selective and stable type of receiving circuit is employed. Methods and apparatus have been developed for measuring the strength of the electromagnetic field which is delivered to the receiving point and for measuring the interference produced by static.

The transmission tests so far have been conducted on a wave length of 5260 meters (57,000 cycles per second). The results of the measurements during the first quarter of the year on the transmission from the United States to England show large diurnal variations in the strength of the received signal and in the radio noise strength, as is to be expected, and correspondingly large diurnal variations in the ratio of the signal to noise strength and in the resulting reception of spoken words. Also, the measurements, although as yet incomplete, show a large seasonal variation.

The character of the diurnal and seasonal variations is clearly indicated in the figures. The curves present the most accurate and complete data of this kind yet obtained.

ON January 15, of this year, a group of about 60 people gathered in London at a prearranged time and listened to messages spoken by officials of the American Telephone and Telegraph Company from their offices at 195 Broadway, New York City. The transmission was conducted through a period of about two hours, and during this time the words were received in London with as much clearness and uniformity as they would be received over an ordinary wire telephone circuit. During a part of the time a loud speaker

[1]This paper, with the exception of the Appendix, was presented at the Annual Convention of the A. I. E. E., Swampscott, Mass., June 26-29, 1923, and was printed in the Journal for August, 1923.

was used in connection with the receiving set, instead of head receivers. The reporters present easily made a transcription of all the remarks, both with head sets and with the loud speaker.

These tests were made possible by cooperation between the engineers of the American Telephone and Telegraph Company and the Western Electric Company, and the engineers of the Radio Corporation of America and its associated companies. The sending apparatus was installed in the station of the Radio Corporation of America, at Rocky Point, L. I., in order to make use of that company's very efficient multiple-tuned antenna. The receiving apparatus was installed in the buildings of the Western Electric Company, Ltd., at New Southgate, England.

This was not the first time speech had been transmitted from America to Europe. Transatlantic telephony was first accomplished in 1915, when the American Telephone and Telegraph Company transmitted from the Navy station at Arlington, Va., to the Eiffel Tower, Paris. In these earlier tests, however, speech was received in Paris only at occasional moments when transmission conditions were exceptionally favorable. The success of the present tests indicates the large amount of development which has been carried out since this first date.

The recent talking tests were carried out as part of an investigation of transatlantic radio telephony. This investigation is directed at determining (1) the effectiveness of new methods and apparatus which have been developed for telephonically modulating and transmitting the large amounts of power necessary for transoceanic operation, (2) the efficacy of improved methods for the reception of this transmission and for so selecting it as to give an extremely sharp differentiation between the range of frequencies transmitted and all the frequencies outside of this range; and (3) determining the transmission characteristics for transatlantic distances and the variation of the characteristics with the time of day and the season of the year, including the measurement of the amount of static interference.

The tests are being continued, particularly as regards the study of transmission efficiency.

Single Side-Band Eliminated Carrier Method of Transmission

The method of transmission used in these experiments is what we know as the single side-band eliminated carrier method[2]. With this

[2]For a more complete exposition of this method see U. S. patent No. 1449382 issued to John R. Carson to whom belongs the credit for having first suggested it. Also see Carson patents Nos. 1,343,306 and 1,343,307.

method, the narrowest possible band of wave lengths in the ether is used, and all of the energy radiated has maximum effectiveness in transmitting the message.

As had been pointed out in other papers[3], when a carrier is modulated by telephone waves, the power given out is distributed over a frequency range, and may be conveniently considered in three parts: (1) energy at the carrier frequency itself, (2) energy distributed in a frequency band extending from the carrier upward, and having a width equal to the frequencies appearing in the telephone waves, and (3) energy in a band extending from the carrier downward, and having a similar width. The power at the carrier frequency itself makes up somewhat more than two-thirds of the total power, even when modulation is as complete as possible. Furthermore, this energy can, in itself, convey no message, as is self evident. In the present method, therefore, the carrier-frequency component is eliminated, by methods explained in detail below with the result that a large saving in power is effected. Each of the remaining frequency ranges, generally known as the upper and the lower side-band respectively, transmits power representing the complete message. It is therefore unnecessary to transmit both of these side-bands, so that in the present method one of them is eliminated. In this way the transmission of the message uses only half the frequency band required in the usual method of operation. Similarly the frequency-band accepted by the receiving set is narrowed to conform to a single side-band as compared with the usual double side-band reception, and as a result the ratio of signal to interference is improved. Certain other advantages of this method will be brought out in the further discussion.

While these advantages of the single side-band eliminated carrier method hold good for radio telephone transmission generally, they become of the utmost importance in transoceanic work, because of the necessity of conserving power in a system where the transmitting powers are large, and also because the very limited frequency range available for long distance transmission makes it imperative that each part of the range shall be utilized with the greatest of care. Before discussing the method further, the circuits and apparatus which are actually used in the tests will be described.

[3] "Carrier Current Telephony and Telegraphy" by Colpitts and Blackwell. Journal A. I. E. E., April, 1921.

"Application to Radio of Wire Transmission Engineering" by Lloyd Espenschied. Proc. Inst. Radio Engrs., Oct. 1922.

"Relations of Carrier and Side-bands in Radio Transmission" by R. V. L. Hartley. Proc. Inst. Radio Engrs., Feb. 1923.

The Transmitting System

The transmitting system is shown in simplified circuit form in Fig. 1. It is illustrated as grouped into three parts: The low-power modulating and amplifying stages, shown below in light lines; the high-power amplifiers, shown in heavy lines above and to the right; and the rectifier which supplies the power amplifier with high-tension direct current, shown in the upper left-hand portion of the diagram.

Referring first to the low-power portion of the system, it will be seen that the voice currents (from either a telephone line or a local microphone) are fed into a balanced type of modulator circuit and are modulated with a carrier current of a frequency of about 33,000 cycles. The operation of the balanced type of modulator in suppressing the unmodulated carrier component is explained in the Colpitts and Blackwell carrier current paper referred to above. The result of this modulating action is to produce in the output circuit of modulator No. 1, modulated current representing the two side-bands, for example, the upper one extending from 33,300 to 36,000 cycles and the lower one from 32,700 down to 30,000 cycles. These components are impressed upon a band filter circuit which selects the lower side-band to the exclusion of the upper one and of any remaining part of the carrier, with the result that only one side-band is impressed upon the input of the second modulator. This second modulator is provided with an oscillator which supplies a carrier current of 88,500 cycles. The result of modulation between the single side-band and this carrier current is to produce a pair of side-bands which are widely separated in frequency, the upper one, representing the sum of the two frequencies, extending from 118,500 to 121,200 cycles and the lower one, representing the difference between the two frequencies, extending from 58,500 down to 55,800 cycles. In this second stage of modulation there is a relatively wide separation between the two-side bands which facilitates the selection at these higher frequencies of one side-band to the exclusion of the other. Another important advantage is that it allows a range of adjustment of the transmitted frequency without changing filters. This is accomplished by varying the frequency of the oscillator in the second step. In the present case, the frequency desired for transmission is that corresponding to the lower side-band of the second modulator. The lower side-band of from 58,500 to 55,800 is therefore selected by means of the filter indicated. This filter excludes not only the other side-band but also any small residual of 90,000-cycle un-

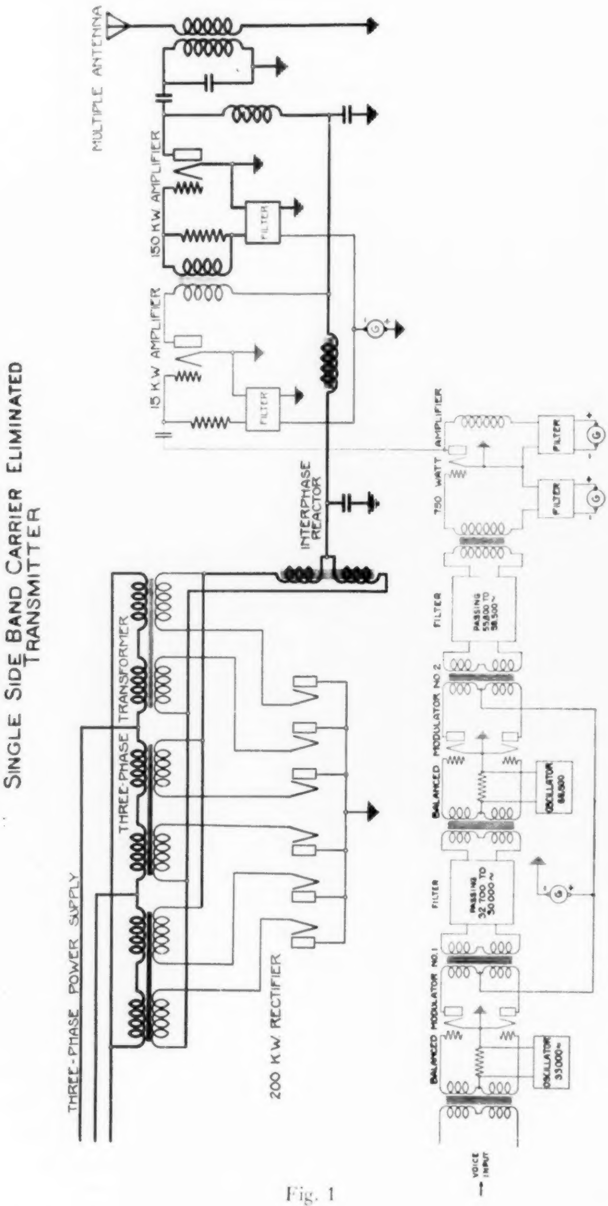

SINGLE SIDE BAND CARRIER ELIMINATED TRANSMITTER
MULTIPLE ANTENNA
150 KW AMPLIFIER
15 KW AMPLIFIER
FILTER
INTERPHASE REACTOR
THREE-PHASE POWER SUPPLY
THREE-PHASE TRANSFORMER
200 KW RECTIFIER
750 WATT AMPLIFIER
FILTER
BALANCED MODULATOR NO. 2
FILTER
BALANCED MODULATOR NO. 1
VOICE INPUT

Fig. 1

modulated carrier current which may get through the second modulator circuit if it is imperfectly balanced.

Having prepared at low power the side-band currents of desired frequency it is necessary to amplify them to the required magnitude for application to the transmitting antenna. This amplification is carried out in three stages. The first stage increases the power to about 750 watts, and is shown in Fig. 1 together with the modulating circuits. This amplifier employs in its last stage three glass vacuum tubes rated at 250 watts each and operating at 1500 volts.

The output of the 750 watt amplifier is applied to the input of the larger-power amplifying system beginning with the 15-kw. amplifier of Fig. 1. This consists of two water-cooled tubes in parallel, operating at approximately 10,000 volts. The output of this amplifier is applied by means of a transformer to the input of the 150-kw. amplifier which consists of two units of ten water-cooled tubes each, all operating in parallel at about 10,000 volts.

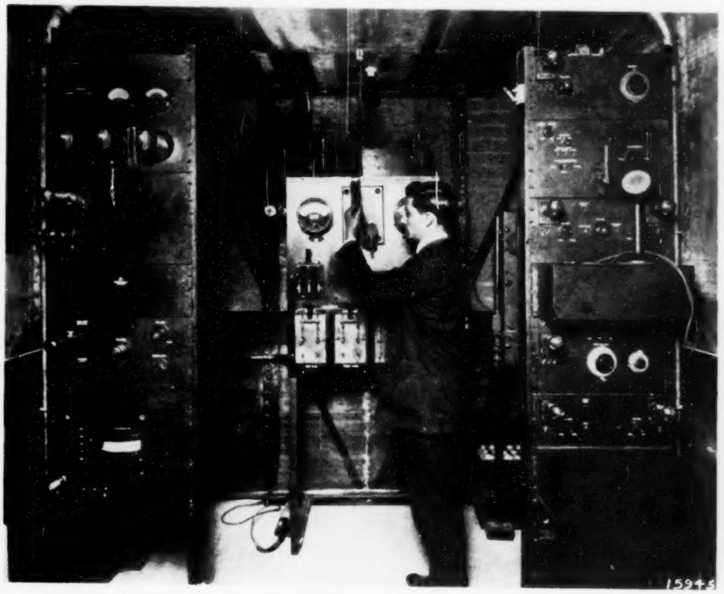


Fig. 2

The high-voltage, d-c. supply is furnished by a large vacuum tube rectifier unit rated at 200 kw. It employs water-cooled tubes similar

to those used in the power amplifiers except that they are of the two-electrode type. The rectifier operates from a 60-cycle, three-phase supply circuit and utilizes both halves of each wave. The two sets of rectified waves are combined by means of an inter-phase reactor which serves to smooth out the resultant current and by distributing the load between tubes of adjacent phases increases the effective load capacity of the rectifier. The ripple is further reduced by the filtering retardation coil and condensers shown.

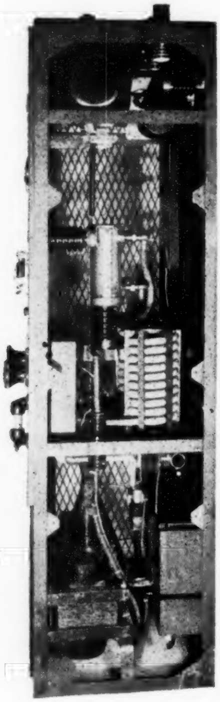


Fig. 3

Reproductions of the apparatus comprising the transmitter system as described above are given in Figs. 2, 3, 4 and 5.

Fig. 2 shows the apparatus comprising the low-power stage of the transmitting system. The right-hand rack contains the two weak-power modulating units and the two single-side-band selecting filters. The left-hand rack is the 750-watt amplifier unit. The three radiation-cooled tubes of 250-watt capacity each will be seen near the top.

Below are the smaller amplifying stages. The power supply board is shown in the center of the photograph.

Fig. 3 is a side view of the 15-kw. amplifier unit. The face of the panel from which the control handles protrude is on the left. Mounted in the cage behind the front panel are two water-jackets for accommodating the water-cooled tubes, also a coiled hose for increasing the electrical resistance of the water supply circuit (the water-cooled anodes of the tubes being operated above ground potential).

The final amplifier of 150-kw. capacity is shown in Fig. 4. It comprises two units each of 75 kw. Each unit contains 10 water cooled tubes which can be seen mounted in their water jackets. To the right of these units is located the 200-kw. rectifier unit shown in Fig. 5. The unit contains actually 12 tubes, there being two tubes

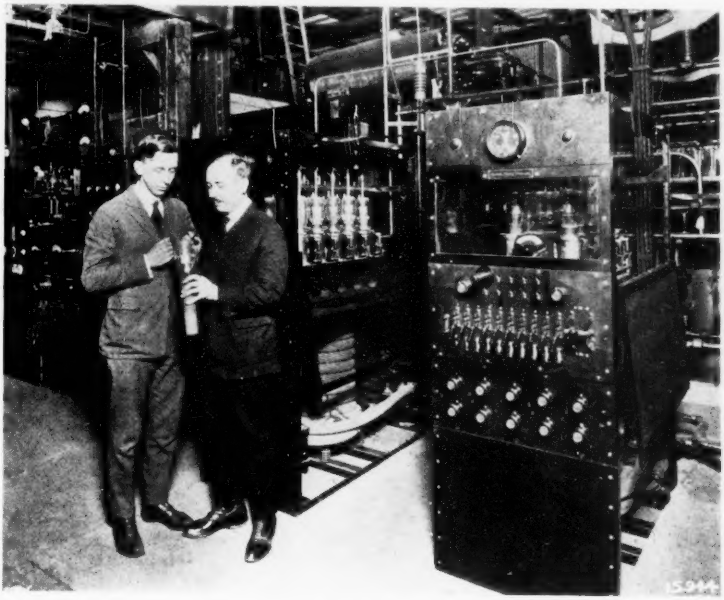


Fig. 4

for each of the six half waves. The pancake coils on the top of the rack are protecting choke coils to guard the transformer secondary winding against steep wave fronts in case of tube failure.

From the above description it will be understood that the transmitting system is one in which the useful side-band is first developed

by modulation and filtration at low power and then its power is built up to a large value in a succession of powerful amplifiers. It

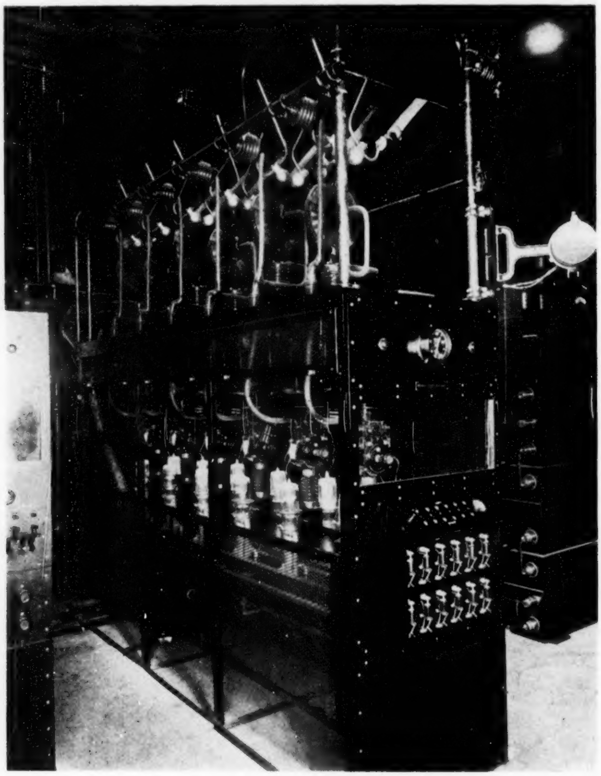


Fig. 5

will be appreciated, therefore, that the large-power amplifiers and in particular the water-cooled tubes which are their essential elements represent one of the major problems of the development.

High-Power Tubes

The development of the high-power tubes is described quite fully in another paper[4]. The present discussion is, therefore, limited to a few of the outstanding features.

[4] *Bell System Technical Journal*, July 1922

In the design of high-power tubes for use in this system the main problem is to insure the ready disposal of the large amounts of heat generated at the anodes. For the conditions of use in the present type of system where the tube is employed as an amplifier, the power which must be disposed of as heat at the anode is of the same order of magnitude as the power which the tube will deliver to the antenna. In the case of the present equipment, therefore, the tube must be so designed as to operate continuously with a heat dissipation at the anode of more than 10 kw. It is obviously difficult to secure so large a dissipation in a tube enclosed with glass walls, and a tube was therefore designed in which the anode forms a part of the wall of the containing vessel and the heat generated in it is removed by circulating water. The tube used is shown in Fig. 6. The lower cylindrical portion is the anode which is drawn from a sheet of copper. The

Fig. 6

upper portion is of glass and serves both to support and insulate the grid and filament elements.

The three principal difficulties met in the construction of these tubes are the making of a vacuum-tight seal between the copper and the glass, the provision of adequate means for conducting through the glass wall the large currents necessary to heat the filament, and the obtaining of the necessary vacuum for high-power operation.

The first of these problems was solved by the development of a new metal to glass seal. In making this seal the glass and metal parts are brought into contact while hot, the temperature being high enough for the glass to wet the metal. The part of the metal in contact with the glass is made so thin that the stresses which are set up when the seal cools are not great enough to fracture the glass or to break it away from the metal at the surface of contact. Seals made in this way are sufficiently rugged to stand repeated heating and cooling from the temperature of liquid air to that of molten glass without deterioration.

A seal employing the same principle but different in form is also used at the point where the leads carrying the filament current pass through the glass walls of the tube. The lead is made of copper 0.064 in. in diameter and passes through the center of a copper disk, 0.010 in. thick, the joint between the lead and the disk being made vacuum-tight by the use of a high melting solder. The disk is sealed to the end of a glass tube which is in turn sealed into the glass wall of the vacuum tube.

In exhausting the tubes it has been found necessary to subject all the metal parts to a preliminary heat treatment in a vacuum furnace during which the great bulk of the occluded gasses is removed. By this method the time of exhaust can be considerably reduced but the vacuum conditions to be met are so stringent that the final processes of evacuation must be carefully controlled and often occupy as much as twelve hours.

The tubes are operated at a plate voltage of 10,000 volts and are capable of delivering 10 kw. at this voltage in a suitable oscillatory circuit. For this performance an average electron current of 1.35 amperes is required. The total electron current that the filament must be capable of supplying to insure steady operation is about 6 amperes.

When the tubes are used to amplify modulated currents with large peak values such as are characteristic of telephone signals it is essential that the maximum electron current through the tube shall be several times the normal operating current and therefore to insure the necessary high quality of transmission these tubes are operated for telephone purposes with an average output of about 5 kw.

The Receiving System

In the method of transmission ordinarily employed in radio telephony by which the carrier and both side-bands are sent out from the transmitting station and received at the distant end, detection is readily

accomplished merely by permitting all of these components to pass through the detector tube. The detecting action whereby the voice-frequency currents are derived, is accomplished by a remodulation of the carrier with each side-band.

With the present eliminated carrier method of transmission the side-band is unaccompanied by any carrier with which to remodulate in the receiving detector. It is necessary, therefore, to supply the

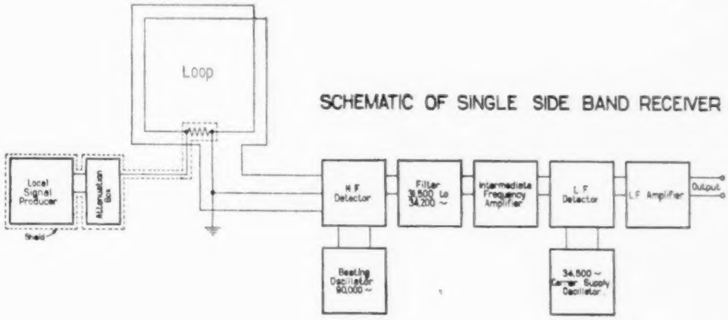


Fig. 7

detector with current of the carrier frequency obtained from a local source. Thus, in the present experiments, if a current of the original carrier frequency, 55,500 cycles, is supplied to the detector it will remodulate or "beat" with the received side-band of, say 55,800 to 58,500 cycles and a difference-frequency band of 300 to 3000 cycles, *i.e.*, the voice frequency band will result.

The arrangement actually used, however, is not quite so simple as this. It is shown schematically in Fig. 7. Reception is carried out in two steps, the received side-band being stepped down to a lower frequency before it is detected. The stepping down action is accomplished by combining in the first detector the incoming band of 55,800 to 58,500 cycles with a locally generated current of about 90,000 cycles. In the output circuit of the detector the difference-frequency band of 34,200 to 31,500 cycles is selected by a band filter and passed through amplifiers and thence to the second detector. This detector is supplied with a carrier of 34,500 cycles which, upon "beating" with the selected band, gives in the output of the detector the original voice-frequency band.

The object of thus stepping down the received frequency is to secure the combination of a high degree of selectivity with flexibility in tuning. The high selectivity is obtained by the use of a band filter.

It is further improved by applying the filter after the frequency is stepped down rather than before. To illustrate this improvement assume that there is present an interfering signal at 60,000 cycles, 1,500 cycles off from the edge of the received telephone band. This is a frequency difference of about $2\frac{1}{2}$ per cent; but after each of these frequencies is subtracted from 90,000 cycles, the difference of 1500 cycles becomes almost 5 per cent. This enables the filter to effect a sharper discrimination against the interfering signal. Furthermore, the filter is not required to be of variable frequency as would be the

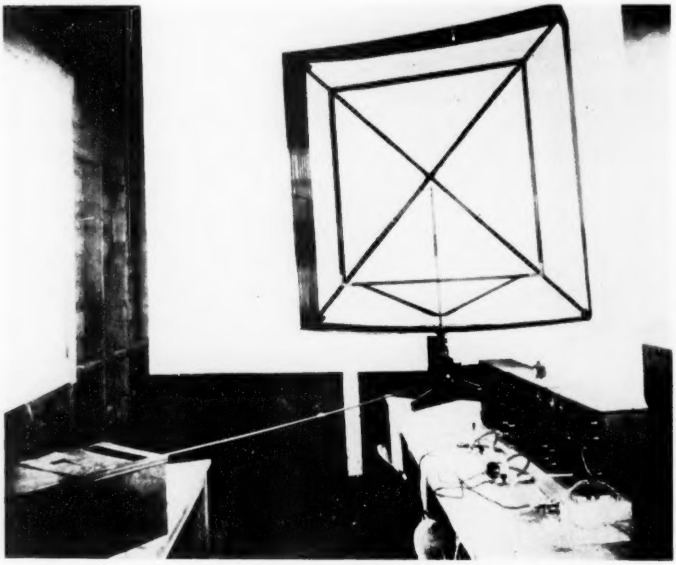


Fig. 8

case were it employed directly at the received frequency since by adjusting the frequency of the beating down oscillator the filter is in effect readily applied anywhere in a wide range of received frequencies. The receiving method, therefore, enables the filter circuit, and indeed also the intermediate frequency amplifiers, to be designed for maximum efficiency at fixed frequency values without sacrificing the frequency flexibility of the receiving set.

A photograph of the receiving set used in the transatlantic measurements is reproduced in Fig. 8. The signals are received on a square

loop six feet on a side and wound with 46 turns of stranded wire. The first box contains the beating oscillator and high-frequency detector, the second box of the filter and amplifying apparatus for the intermediate frequency and the third box the final detector and amplifier. The shielded box at the left of the picture, which is connected to the loop by means of leads in the copper tube, is the apparatus for introducing the comparison signal of known strength into the loop for measuring purposes. This receiving and measuring set is described more in detail in a paper by Bown, Englund, and Friis in the "Proceedings of the Institute of Radio Engineers for April, 1923."

Although it was this very selective and reliable method of intermediate-frequency reception which was used in London, it is quite possible to receive the single-side-band transmission by means of a regular heterodyne receiving set. Even a self-regenerative set will suffice under some conditions. It is necessary, however, to adjust the frequency of the oscillator very carefully to that of the transmitting carrier frequency, otherwise serious distortion of the received speech will result. Also it is, of course, necessary that the tuning be not too sharp if ordinary tuned circuits and not filter circuits are employed. One might expect that some difficulty would be experienced in maintaining the frequency at the receiving end in sufficiently close agreement with the sending frequency. In the tests no particular difficulty was experienced, the oscillators at the two ends being so stable that only an occasional slight readjustment of the receiving oscillator was required. With the development of more stable oscillators, doubtless, the frequency with which readjustments are required, will be further reduced. If serious distortion of the received speech is to be avoided the two frequencies must be within about 50 cycles, an accuracy of 0.1 per cent at 50,000 cycles.

Transmission Advantages of the System

Since the present experiment represents the first use of the single-side-band eliminated carrier type of system some further discussion of the characteristics and advantages of the system is appropriate.

The importance of the system in conserving frequency range will be appreciated when it is realized that the total frequency range available for transatlantic telephony is distinctly limited. Just what the most suitable range is has not been accurately determined but it seems limited to below 60,000 cycles (5000 meters) because of the large attenuation suffered during the daylight hours by frequencies higher than this. On the lower end of the frequency scale, trans-

atlantic telegraphy at present pretty well preempts frequencies below 30,000 cycles (10,000 meters). Therefore, for the present at least transatlantic telephony is limited to a range of some 30,000 cycles. Now transmission of speech requires as a minimum for good quality a single-side-band 3000 cycles wide. Allowing for variations and clearances between channels it is doubtful if the channels could be made to average closer than one every 4000 cycles for single-side-band transmission and one every 7000 cycles for the ordinary double-side-band transmission. This means that even were the whole range from 30,000 to 60,000 cycles devoted to telephony to the exclusion of telegraphy, only about four channels could be obtained by the older methods and some seven by the present one.

It is a rather interesting commentary to note that a somewhat similar situation as to limitation in frequency range exists in the case of carrier-current transmission over wires. The transmission efficiency falls off with increase in frequency and limits the range of frequencies which can be economically used, in much the same way as it is limited in long distance radio transmission. It is because of this limitation in the case of wires and the value which attaches to conserving the frequency range consumed per message that single-side-band transmission was first developed for wire carrier current systems. Its development in wire transmission has been of considerable aid in adapting the method to the present purpose of transatlantic operation.

The second of the outstanding characteristics of the present system resides in the large power economy which it permits. Transatlantic telephony requires hundreds of kilowatts of high-frequency power. Since it is difficult and expensive to produce this power it is important that every effort be made to increase its efficiency or effectiveness in transmitting the voice. To illustrate how the present system effects economies in power, consider the case of a carrier wave completely modulated by a single frequency tone. In such a completely modulated wave, only 1/3 of the total power contains the message, the remaining 2/3 conveying only the carrier frequency which can as well be supplied from an oscillator of small power at the receiving station. It is obvious, therefore, that by eliminating the carrier only 1/3 as much power need be used as would be required were all the elements of the completely modulated wave transmitted. To realize the maximum advantage of this mode of operation, the system eliminates the carrier at low power and, thereby, the high-power apparatus is devoted exclusively to the amplification of the essential part of the signal.

If, after having suppressed the carrier, both side-bands were transmitted, their reception would require perfect synchronism between the carrier resupplied at the receiving end and that eliminated at the sending end, a condition which is practically impossible to meet without transmitting some form of synchronizing channel, which is, indeed, much the same as transmitting the carrier itself. If the receiving carrier is not synchronized, the two side-bands will interfere with each other upon being detected. By eliminating one side-band, this interference is prevented and reception may be carried on, using a locally supplied frequency which is only approximately equal to that of the suppressed carrier. The two may differ by as much as 50 cycles before the quality of the received speech is greatly impaired. The importance to the carrier suppression method of eliminating one side-band will, therefore, be appreciated. The present system eliminates one side-band while still in the low-power stage. While it would be possible to do this selecting after they have both been raised to the full transmitting power, this would require the use of a filter of high-power carrying capacity, which would make the filter very costly and also render the system inflexible to change of wave length. The present system overcomes both of these difficulties by filtering our one side-band at low-power levels and by the use of the double modulation method.

Another very important reason for the transmission of a frequency band as narrow as is possible lies in the difficulty of constructing an antenna to transmit more or less uniformly at these long waves a band of frequencies which is an appreciable fraction of the main carrier frequencies. For example, in the ordinary method of transmission an antenna which was intended to transmit a 30,000-cycle carrier and its two speech side-bands would need to be designed to transmit all the frequencies from 27,000 cycles to 33,000 cycles, a band which is equal to 20 per cent of the carrier frequency. This band is considerably wider than that given by the resonance curve of a highly efficient long wave antenna. To accommodate both side-bands would require flattening out the resonance curve either by damping, which means sacrifice in power efficiency, or by special design of the antenna, possibly throwing it into a series of interacting networks and causing it to become a rather elaborate wave filter. The importance, from the antenna standpoint, of narrowing the frequency band required to be transmitted is, therefore, evident.

It is extremely important that the received signal be affected as little as possible by changes in the transmission efficiency of the medium. The voice frequency currents produced at the receiving

end, after detection, are proportional to the product of the carrier wave and the side-band. If the carrier as well as the side-band is transmitted through the medium, then a given variation in the transmission efficiency of the medium will affect both components and will change the received speech in proportion to the square of the variation, as compared to the first power if only the side-band is transmitted and the carrier is supplied locally. Thus it will be seen that the omission of the carrier from the sending end and the resupplying of it from the constant source at the receiving end gives greater stability of transmission.

Without discussing the system in further detail the advantages of it may be summarized as follows:

1. It conserves the frequency (wave length) band required for radio telephony, which is particularly important at long wave lengths.
2. It conserves power, in that all of the power transmitted is useful signal-producing power. This is particularly important also in long distance transmission which requires the use of large powers.
3. The fact that only a single-band of frequencies is transmitted simplifies the antenna problem at long wave lengths, where the resonance band becomes too narrow to transmit both side-bands.
4. As compared with a system which eliminates the carrier but transmits both side-bands the simple side-band system has the important advantage of not requiring an extreme accuracy of frequency in the carrier which is resupplied at the receiver. Were both side-bands transmitted very perfect synchronism would be required for good quality.
5. It improves the transmission stability of the radio circuit since variations in the ether attenuation affect only one (the side-band) of the two components effective in carrying out the detecting action in the receiver.
6. The receiving part of the overall system has two advantages:

a. It need accept only half of the frequency band which would be required in double side-band transmission, thereby accepting only half of the "static" interfering energy.

b. By stepping down the frequency of the received currents and filtering and amplifying at the low-frequency stage a very sharp cutoff is obtained for frequencies outside of the desired band and a very stable and easily maintained amplifying system is obtained.

Study of Transatlantic Transmission

We come now to a consideration of the second major part of the investigation, namely, that having to do with the transmission of the waves across the Atlantic. It will be evident, from what has been said earlier, that the transmission question is essentially one of how best to deliver, through the variable conditions of the ether to the receiving station, speech-carrying waves sufficiently free from interference to be readily interpretable in the receiving telephone. The transmission efficiency of the medium varies with time of day and year, and is different for different wave lengths. The interference conditions are also influenced by these same factors.

Now we can study this transmission medium in much the same way we would a physical telephone circuit, by putting into it, at the sending end, electromagnetic waves of a known amount of power and measuring the power delivered at the receiving end. The interference at the receiving station likewise may be measured and the ratio of the strength of the signal waves to the interfering waves may be taken as a measure of freedom from interference; this in turn being directly related to the readiness with which the messages are understood. Accordingly, there has been included as an integral part of the investigation of transatlantic radio telephony, the development of suitable methods and apparatus for measuring the strength of the signal waves and of the interfering waves, as they arrive at the receiving station. The apparatus[5] employed in measuring the field strength of the received signals has been outlined above under Receiving System and need not be gone into further. However, a word of explanation about the method which is employed in making the measurement may be helpful. It will be recalled that the specially designed receiving set is provided with a local source of high frequency from which can be originated signals of predetermined strength. A measurement of the field strength of a signal received from the distant transmitter is made by listening first to the distant signal and then to the locally produced signal, shifting back and forth between these signals and adjusting the strength of the local signal until the two are substantially of the same strength. Then, knowing the power delivered by the local source, the power received from the distant station is likewise known. The relation between the power in the input of the radio receiving circuit to the field strength required to deliver that power is known through the geometry of the receiving

[5]It is described in detail in the paper entitled, "Radio Transmission Measurements" by Bown, Englund, and Friis, *Proc. Institute of Radio Engrs.*, April, 1923.

antenna (in this case a loop) and, therefore, the measured power of the signal can be translated directly into the field strength of the received waves.

The measurement tone signal is transmitted from the Rocky Point sending station by substituting for the microphone telephone transmitter a source of weak alternating current of about 1/100 watt at a frequency of approximately 1500 cycles. This tone modulates the radio telephone transmitter in the same way that voice currents would and is radiated from the antenna as a single-frequency wave of 5260 meters (57,000 cycles per second). It, therefore, constitutes a means of sending out a single-frequency continuous wave for measurement purposes. At the receiving end this continuous wave is demodulated to the same tone frequency which it originally had.

For measuring the strength of the received noise, *i.e.*, the radio frequency currents arising from static or other station interference, the method is quite similar. In this case, however, the noise received is so different from that which can be set up artificially in any simple manner that no attempt is made to compare it directly with a local noise standard. Instead the volume of the interfering noise is expressed in terms of its effect in interfering with the audibility of a local tone signal by measuring the local signal which can just be definitely discerned through it. This is a threshold type of measurement which is necessarily difficult to carry out with accuracy. In order to increase the sharpness of definition of the local signal and to make it correspond more closely to speech reception the signal tone is subjected to a continuous frequency fluctuation. The comparison signal has therefore a warbling tone which occupies a frequency band not unlike that of the voice. This method of measuring the interference is discussed in more detail also in the measurement paper referred to above.

Procedure in Making Transmission Measurements. The three quantities which are included in the transmission measurements, namely, the signal strength, the noise strength, and the percentage of words received correctly, are observed one after another in what might termed a unit test period. Although the duration of this test period and the order of making the measurements has been changed somewhat during the course of the experiments, the following program is representative of the conditions under which the data presented below were taken.

A 25-minute test period divided as follows:

5 minutes of tone telegraph identification signals (for receiving adjustment purposes).

10 minutes of disconnected spoken words.

10 minutes of a succession of five-second tone dashes separated by five-second intervals, (for measurement of the received field strength, the intervals between the dashes being used for throwing on the local receiving source and adjusting its strength to equal that of the receive signals by alternately listening to one and then the other).

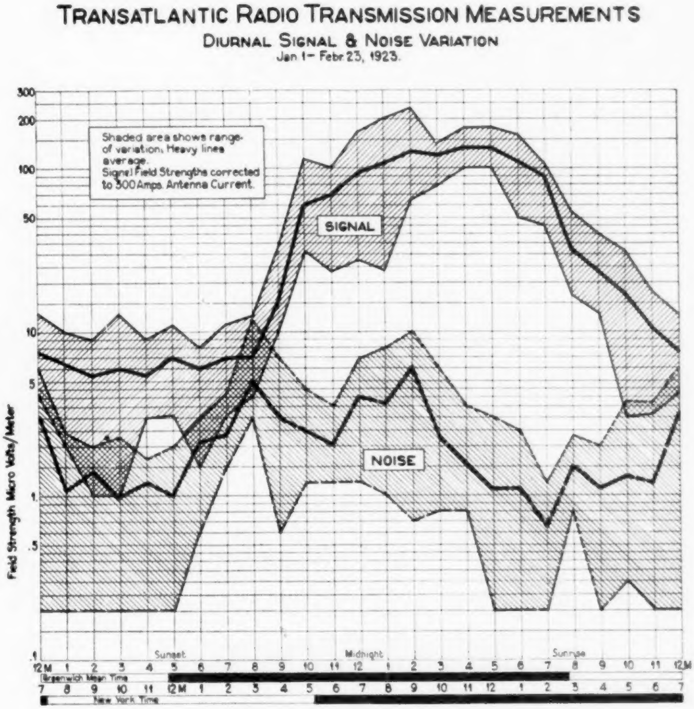


Fig. 9

Immediately following this test period the London observers measured the noise level.

This unit test period was repeated every hour over a period which varied from several hours to as long as two days' duration. Most of the test periods ran for about 28 hours, starting about eleven o'clock Sunday morning and continuing until about three o'clock Monday morning, London time. During this time the telegraph load through

the Rocky Point station of the Radio Corporation was sufficiently light to enable one of the two antennas to be devoted to these experiments. The measurements were started January 1, 1923 and are still in progress.

At the present time (April) the results for the first three months of the tests are available. These results are not yet sufficiently complete nor do they cover a sufficient number of variables in terms

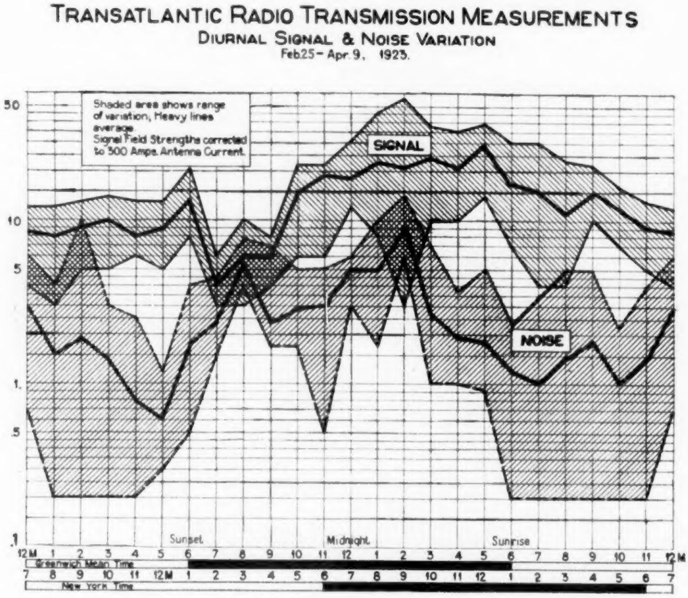


Fig. 10

of time, wave length, etc., to enable any very definite conclusions to be drawn from them. They do illustrate, however, the usefulness of the methods employed, and even in their incomplete state show some factors of considerable interest.

The results of the measurements of received signal, strength and received noise are given in Figs. 9 and 10. The data have been divided and plotted in these two sets of curves because the transmission conditions across the North Atlantic appeared to suffer a rather rapid change about February 23rd. Fig. 9 therefore covers the winter period from January 1 (when the test started) to February 23; and Fig. 10 covers the next period from February 25 to April 9.

The curves are plotted between time of day as abscissas and field strength in microvolts per meter as ordinates. The time during which darkness prevailed at Rocky Point and at London is indicated by the block-fills on the time scales. The overlap of these block-fills indicates the time during which darkness extended over the entire transatlantic path. For Fig. 9 the darkness-belt is as of February 1 and for Fig. 10 as of March 21. The curves show the mean of the results and also the boundaries of the maximum and minimum values observed.

Received Signal Strength. The outstanding factors to be noted concerning the signal strength curves are:

1. The diurnal variations are plainly in evidence. During the first test period covered by Fig. 9, for example, the field strength varied in the ratio of the order of 15 to 1 between day and night conditions, running about 100 microvolts per meter during the night and averaging about 6 microvolts per meter during the day. The diurnal variation is also to be seen in Fig. 10 although the variations between night and day transmission are less marked.

The measured daylight values lend support to the Austin-Cohen absorption coefficient. The average of the observed daylight value for the period of these tests is between 7 and 8 microvolts per meter, while the calculated value is 9.5. Concerning the high field strengths obtaining at night, it should be noted that the maximum observed value of 237 microvolts per meter does not exceed the value of some 340 microvolts which it is estimated should obtain at London were no absorption present in the intervening medium, *i.e.*, were the waves attenuated in accordance with the simple inverse-with-distance law. While no definite conclusions can yet be drawn from these results as to the cause of the diurnal variations, this indication that the upper limit of the variation is the no-absorption condition suggests that the diurnal fluctuations are controlled by the absorption conditions of the medium rather than by reflection or refraction effects.

2. An indication of the seasonal variation which apparently occurs in developing from winter to early spring is found in a comparison of the signal strength curves of Figs. 9 and 10. On the whole the signal strength received in the second test period is considerably less than that received for the first period. This drop in the average of the 24 hours is caused by a large decrease in the night-time transmission efficiency. The daylight transmission does not change much, but what little change there is lies in the direction of an increase as the season advances.

3. A decrease in the transmission efficiency is observed between the time of sundown in London and sundown in New York, that is,

during the period when the sunset condition intervenes in the transmission path. This dip is particularly noticeable in the signal strength curve of Fig. 10. It is not noticeable in Fig. 9, except for the fact that the rise in signal strength corresponding to night conditions in London is delayed until the major part of the transmission path is in darkness.

Strength of Received Noise. The variation in the strength of received noise is shown by the noise curves of Figs. 9 and 10.

1. The diurnal variation of that portion of the noise which is due to atmospheric or "static" disturbances is somewhat obscured by the presence of artificial noise, *i.e.*, noise caused by interference from other stations. The rise in the noise curve at 12 noon is known to be due to artificial interference. In general, however, the large noise values shown to prevail throughout the night in London between about 6 p. m. and 4 a. m. are known to be due to atmospherics. This diurnal variation shows up quite prominently in both figures.

The maximum noise is reached at 2 a. m. London time. Up to this time the night belt extends over London and a sector of the earth considerably to the east and including Europe, Africa and Asia. The noise begins to drop off shortly thereafter and reaches its minimum at sunrise in London. This could be accounted for on the assumption that the major source of the noise lies considerably to the east of London and that transmission of the stray electric waves to London is gradually diminished in efficiency as daylight overtakes the path of transmission.

2. The seasonal variation, as shown by a comparison of the noise curve of Fig. 9 with that of Fig. 10, is not so great as is the case with the transmission efficiency of the signal. However, the noise level is noticeably higher during the second period of the tests, as shown by the average curve of Fig. 10, particularly during the night when the maximum noise obtains.

This indicates that the noise is largely of continental origin lying to the east or south east of London which is in agreement with rough observations made by means of a loop and suggests that the employment of directional antennas would be of considerable advantage. It is expected to include such antennas in the further measurement work.

In connection with these noise curves it should be noted that what they represent is in reality the strength of a local warbling tone-signal, expressed in terms of equivalent field strength in microvolts, which which is just definitely audible through the noise. The actual value of the noise currents, were they measured by an integrating device such as a thermocouple, for example, would be a number of times larger than indicated.

Ratio of Signal to Noise Strength; Words Received. The noise curve of Fig. 9 and that of Fig. 10 can, therefore, be read as "The strength of the signal tone which can just be heard through the noise." It can, therefore, be directly compared with the signal curve itself and the difference between the two curves is a measure of the level of the actual signal strength above that which would just permit of the signals being heard. Actually, the difference between the two curves, as shown in the figures, is proportional to the *ratio* of the signal to the noise strength, because the curves are plotted to a logarithmic scale.

This signal to noise ratio is plotted in Fig. 11 for the test period which corresponds to Fig. 9, and Fig. 12 for the test period which corresponds to Fig. 10. These ratio curves are derived by going back to the original data and taking the ratio for each unit measurement period and spotting it upon the chart as shown by the black points. An average is taken of the points for each hour of the 24-hour period as shown by the circle points. The dash line curves of Figs. 11 and 12, therefore, trace the average diurnal variation of signal to noise ratio.

These curves show:

1. That the signal to noise ratio reaches its minimum during the time when the sunset period intervenes between London and New York.

2. During the night in London the ratio increases more or less continuously and reaches a maximum around the time of sunrise in London.

3. During the course of the daylight period in London the ratio starts out high and drops rather rapidly during the forenoon and assumes a more or less constant intermediate value during the afternoon until sundown. It is during this afternoon period in London that the business hours of the day in London and New York coincide, so that this is the most important period from a telephone communication standpoint.

The drop in the very low ratios obtaining in London in the early evening is due to the fact that an increase in noise occurring at this time is accompanied by a decrease in transmission efficiency from America. This may readily be seen by referring to Fig. 10. The noise increases as the night belt, proceeding westward, envelops England and improves the transmission of atmospherics, which arise possibly in continental Europe, Asia and Africa. As the shadow wall, proceeding westward, intervenes between England and America, the transmission efficiency of the desired signals from America drops and it is not until the night belt extends as far west as America that the transmission efficiency improves sufficiently to overcome the dis-

advantage in London of the large noise values which night there had brought on. Conversely, the high signal to noise ratio, obtaining at about sunrise in London, appears to be due to the fact that as the termination of the night belt, moving westward, intervenes between England and the source of atmospherics to the east, the noise level drops rapidly and has reached low values by the time sunrise arrives in London. At this time, however, darkness still extends to the west

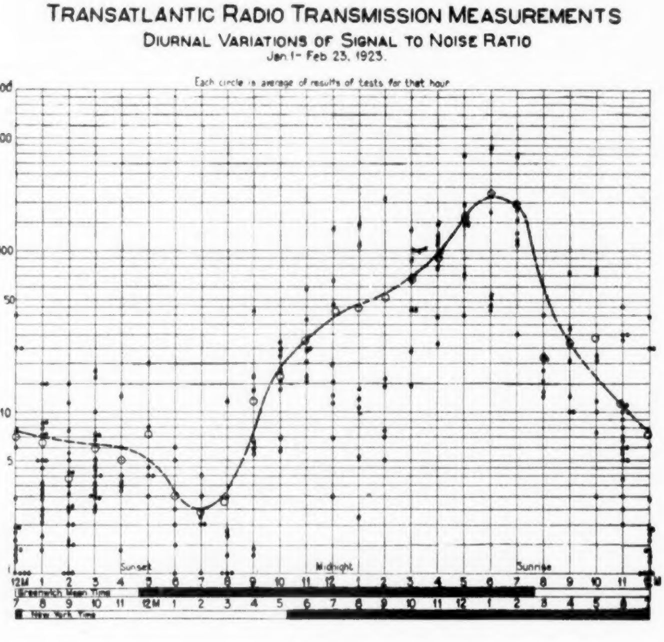


Fig. 11

and the transmission efficiency from America is at its maximum. It is, therefore, due to this interplay between these two factors, signal strength and noise strength, controlled very largely by the transition periods between day and night, that the signal to static ratio varies diurnally in the manner pictured in Figs. 11 and 12.

Concerning seasonal variation, shown by a comparison of Figs. 11 and 12, the following can be said: The dimunition in signal-to-noise ratio in the second test period as compared with the first is caused by the fact that the signal strength has decreased and at the same time the noise has somewhat increased. There is just one other

point that concerns the dip in the ratio occurring at night in London between 12 midnight and 3 a. m. This dip is due to an increase in the noise which occurs around 2 a. m. (A further reduction during this

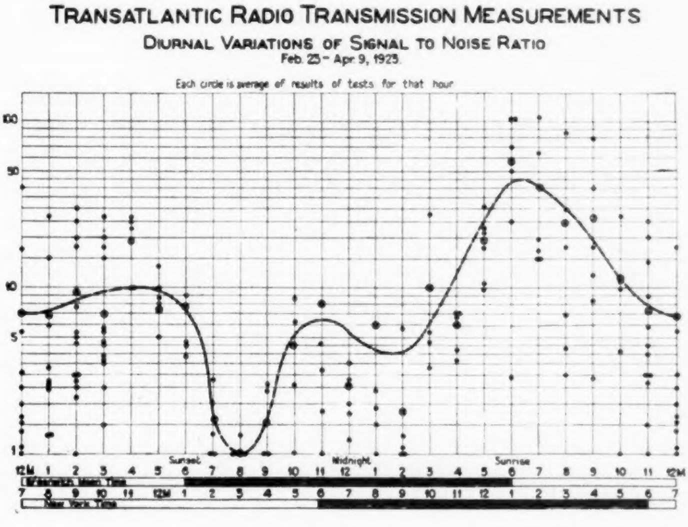


Fig. 12

time, and one which extends the time of minimum ratio from sundown on through the night until 2 a. m. is shown by the April measurements which time has not permitted including in the curves).

During each test period lists of disconnected words were spoken over the systems. As an approximate and easily applied method of indicating the talking efficiency of the circuit, note was made of the percentage of the words which were correctly received.

The curves of Figs. 13 and 14 show the manner in which the percentage of the words which were correctly received varies through the 24 hours. Each point corresponds to the percentage of words correctly received during one unit test period. In many of these tests the interference was noted to be caused by radio telegraph stations, and the data in which the interference is of this character, in so far as identified, are indicated by the triangular dots. It will be seen that most of the poor receptions were due to this cause. Especially is this true of tests at 12 noon at which time severe interference from sources local in London was experienced. The circle points are the

average of results for each hour's tests. The dash line curve is a smoothing out curve of these points.

It is interesting to note that these curves of actual word count conform very well in general shape with those of Figs. 11 and 12 which also really measure receptiveness although in a less direct manner. Reception is best during the late night and early morning,

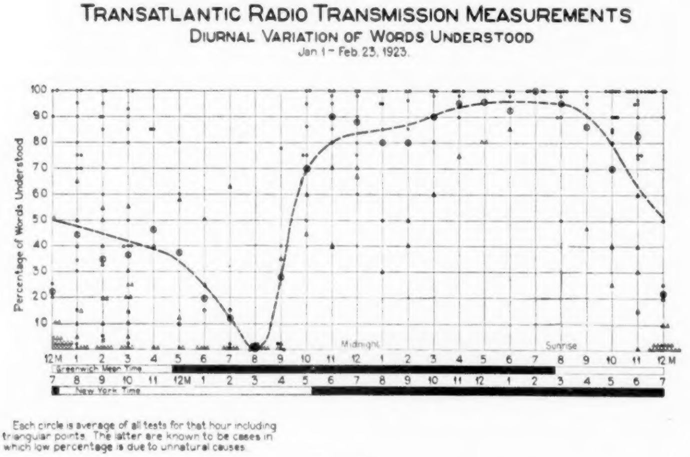


Fig. 13

drops off during the day, reaching a minimum during the evening. Furthermore, the night reception is shown to be considerably better for the January-February period than for the February-March period. The curve of Fig. 14 corresponds quite closely with that of Fig. 12. The curve of Fig. 13 does not show as much of a peak as does that of Fig. 11 which is, of course, due to the fact that above a certain ratio the percentage of words understood is high and cannot rise above 100 per cent.

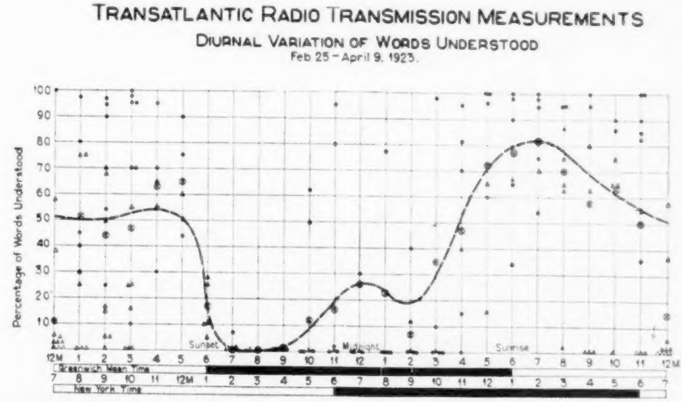
Conclusion

As has been indicated this is a report of work which is still in progress. To date:

A new type of radio telephone system affording important advantages for transatlantic telephony has been developed and put into successful experimental operation across the Atlantic.

Sustained one-way telephonic transmission has been obtained across the Atlantic for the first time by means of this system.

The advantages of this system which had been anticipated, particularly, in respect to economies of power and wave lengths, have been realized. Furthermore, it has been demonstrated that the high-power water-cooled vacuum tubes which have seen their first prolonged operation in this installation are admirably adapted for use in high-power radio installations and particularly for use as high



Each circle is average of all tests for that hour including triangular points. The latter are known to be cases in which low percentage is due to unnatural causes.

Fig. 14

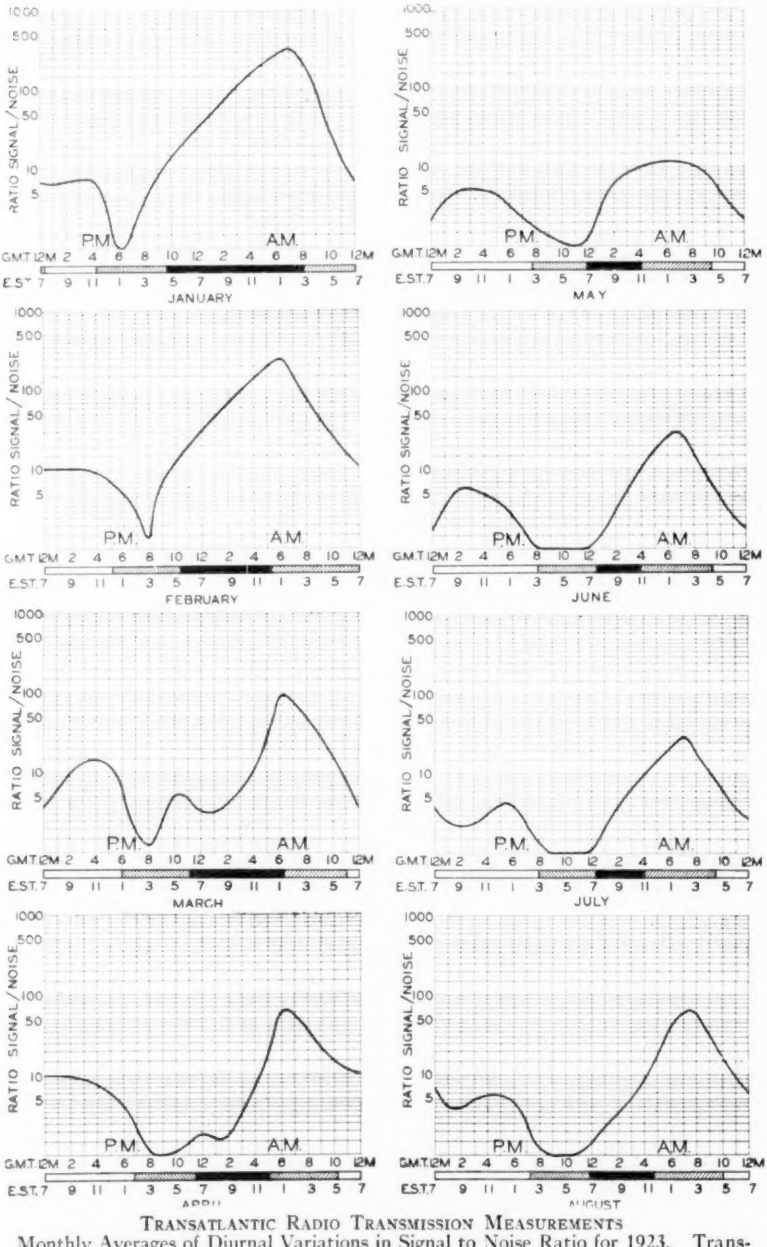
power amplifiers, in the type of system we have described. Also, the method of reception has proved itself to be eminently satisfactory for use with the single side-band type of transmission and to possess important advantages for radio telephony in respect to selectivity and amplification.

Methods have been developed for measuring the strength of the received signals and the strength of the received interfering noise and these methods have been successfully applied in the initiation of a study of the variations to which transatlantic transmission is subject.

The results of the transmission measurements show that, at 5000 meters, the diurnal variations are large, as was to be expected, and give evidences of a large seasonal variation which was, indeed, also to be expected. The results indicate that it will probably be desirable to use a wave length longer than 5000 meters. The measurements are now being made to include the longer wave lengths.

APPENDIX Added September 23, 1923

The results of the transmission measurements from January through August are now available and are summarized in the curves following:



Transatlantic Radio Transmission Measurements
Monthly Averages of Diurnal Variations in Signal to Noise Ratio for 1923. Transmission from Rocky Point to London on 57,000 Cycles (5,260 Meters). Measurements on Loop Reception. Curves Corrected to 300 Amperes Antenna Current

Physical Measurements of Audition and Their Bearing on the Theory of Hearing *

By HARVEY FLETCHER

SYNOPSIS: The author states his purpose to be the presentation of certain facts of audition which have been determined recently with considerable accuracy and the discussion of the theory which best explains these facts.

Making use of data of Knudsen's as well as his own measurements of the auditory sensation area, the author estimates that the normal ear can perceive approximately 300,000 different pure tones. This is taking account of all possible variations in both pitch and intensity. Knudsen's data show that for considerable ranges the minimum perceptible difference in intensity bears a constant ratio to the intensity and the minimum perceptible difference in frequency bears a constant ratio to the frequency. These relations have been termed by psychologists "The Law of Weber and Fechner."

A loudness scale is proposed such that the difference in loudness between two tones is equal to ten times the common logarithm of their intensity ratio. A pitch scale is proposed such that the difference in pitch is equal to one hundred times the logarithm to the base two of the frequency ratio. A method for measuring the loudness of complex sounds is mentioned but is to be discussed in a later paper. A method is proposed for expressing quantitatively different degrees of deafness.

Reference is made to data obtained by the author on the masking of one pure tone by another. The minimum audible intensity of a pure tone depends upon the presence of another tone of different frequency. A low pitched note will, in general, exert a surprisingly large masking effect upon notes of higher frequency. The masking of a low note by a higher is not nearly as pronounced. From his observations, the author draws certain interesting conclusions. For example, given a complex tone consisting of three frequencies 400, 300 and 200 cycles with relative loudness values of 50, 10 and 10, respectively, the ear would hear only the 400 cycle tone and the 200 cycle tone. If the sound is now increased 30 loudness units, without distortion, the 400 cycle tone and the 300 cycle tone only, will be heard.

Binaural masking in which each ear receives one of the two sounds is considered and the conclusion reached that the masking effect noted results from conduction of the masking tone through the bones of the head to the ear receiving the masked tone.

It is stated on the basis of data obtained by Wegel and Lane that the oscillatory system of the ear, comprised by the membranes and little bones of the middle and inner ears, does not obey Hooke's Law regarding the proportionality of stress and strain. Consequently, the ear, when stimulated by a pure tone, introduces harmonics and the workers cited have observed harmonics as high as the 4th order. The non-linear transmission characteristic of the vibratory system of the ear is held to account for the greater masking of a high frequency by a lower.

A theory of hearing is advanced which pictures the basilar membrane as being caused to vibrate by incident sound waves. In the case of a pure tone, the membrane is supposed not to vibrate uniformly throughout its length but the region of maximum amplitude determines the pitch of the tone as interpreted by the ear and the maximum amplitude determines the intensity.—*Editor.*

THE question of how we hear has been a subject for discussion by scientists and philosophers for a long time. Practically every year during the past fifty years articles have appeared discussing the

* Presented at the meeting of the Section of Physics and Chemistry of The Franklin Institute held Thursday, March 29, 1923, and published in the *Journal of the Franklin Institute* for September, 1923.

pros and cons of various theories of hearing. These discussions have been participated in by men from the various branches of science and particularly by the psychologists, physiologists, otologists, and physicists. During the past two or three years this discussion has been particularly acute. It is not uncommon to pick up an article and read in the beginning or concluding paragraphs statements such as the Helmholtz theory of audition seems to have sunk beyond recovery, [90, 65] † and at the same time an article written probably a month later will have the conclusion that the Helmholtz theory of audition is definitely established beyond all controversy.[70-75]

There is apparently a great deal of misunderstanding between various writers because of different points of view due to different training. To the physicist it seems that most of the discussions show a profound ignorance of the dynamics of the transmission of sound by the mechanism of the ear. Those discussions by the physicists are frequently open to criticism by the otologist and psychologist, due to his lack of knowledge of the structure of the ear or the mental reaction involved in the process of interpretation. I think it is fortunate that some of these scientists from the different branches are now cooperating in their research work as is evinced by the appearance of several joint papers. (Papers by Dean and Bunch, Minton and Wilson, Wegel and Fowler, Kranz and Pohlman, and others.)

It is not my purpose to discuss the merits of the various theories of hearing, but I desire to present some of the facts of audition which have been recently determined with considerable accuracy, and then discuss the theory of hearing which best explains these facts.

Hearing is one of the five senses. It is that sense that makes us aware of the presence of physical disturbances called sound waves. For my purpose, sounds may be classified into two groups, namely, pure tones and complex sounds. A pure tone is specified psychologically by two properties, namely, the pitch and the loudness. These sensory properties are directly related to the physical properties, frequency and intensity of vibration. Mixtures of pure tones of different loudness, but of the same pitch, fall under the first class, since such mixtures give rise to a pure tone. The complex sounds are varying mixtures of pure tones. It will be noticed that phase has not been taken into account. Except when using the two ears for locating the direction of sources of sound, phase differences are not ordinarily appreciated by the ear.*

† These numbers refer to the bibliography at the end of the paper.

These tones are usually transmitted by means of air waves through the outer ear canal to the drum of the ear. From here the vibrations are transmitted by means of the bones in the middle ear to the mechanism of the inner ear.

Those facts of audition which are familiar to almost everybody are as follows:

1. Pure tones are sensed by the ear and differentiated by means of the properties *pitch* and *loudness*.

2. When two notes, separated by a musical interval, are sounded together, they are sensed as two separate notes. They would never be taken for a tone having the intermediate pitch. In this respect, hearing is radically different from seeing. When a red and a green light are mixed together, the impression received by the eye is that of yellow, an intermediate color between the two.

3. There is a definite limiting difference in pitch that can just be sensed.[1-9]

4. There is a definite limiting difference in intensity that can just be sensed.[10-14]

5. There is a minimum intensity of sound below which there is no sensation.[15-31]

6. There is an upper limit on the pitch scale above which no auditory sensation is produced.[32-44]

7. There is a lower limit on the pitch scale below which there is no auditory sensation produced.[45-51]

8. The ear perceives tones separated by an octave as being very similar sensations.

Another quality of audition which is not so commonly known was pointed out by A. M. Mayer.[52] He stated that high tones can be completely masked by louder lower tones while intense higher tones cannot obliterate lower ones though the latter are very weak. Experiments to be described later in the paper show that this statement must be modified somewhat. Very intense low ones will produce a masking effect upon still lower tones, although the masking effect is very much more pronounced in the opposite case. Many of the opponents of the Helmholtz resonant theory of hearing claim that this fact is fatal to such a theory.[52]

* This statement may require modification when more experimental data are available. As shown later in the paper the middle ear has a non-linear response. Consequently it would be expected that phase differences, especially between tones which are harmonic, would produce spacial differences in nerve stimulation.

LIMITS OF THE FIELD OF AUDITION

The new tools which have made possible more accurate measurements in audition are the vacuum tube, the thermal receiver and the condenser transmitter. When connected in a proper arrangement of circuits, the vacuum tube is capable of generating an oscillating electrical current of any desired frequency. This electrical vibration is translated into a sound vibration by means of the telephone receiver. Between the receiver and the oscillator, a wire network called an attenuator[28] is interposed which makes it possible to regulate the

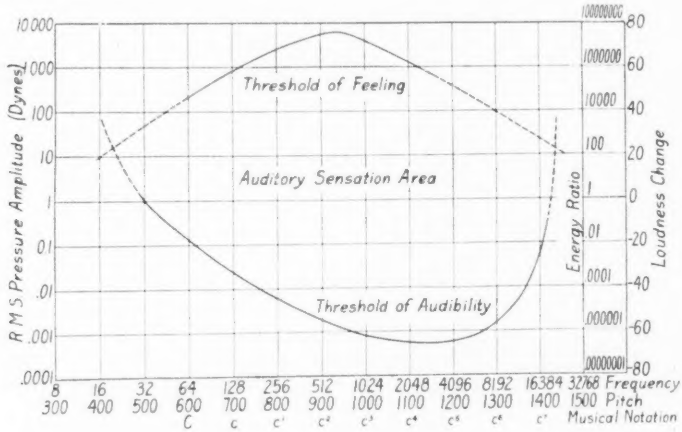


Fig. 1

volume of sound. The theory[54-55] of the thermal receiver has been worked out so that it is possible to calculate its acoustic output from the electrical energy it is absorbing. In this way, it is possible to calculate the pressure variation produced in the outer ear canal when a tone is being perceived. A detailed description of the apparatus and method used in such measurements was given in a paper presented before the National Academy of Science, November 14, 1921.[55] Such a combination of apparatus which has been calibrated is called an audiometer and is suitable for measuring abnormal as well as normal hearing. A receiver more rugged than the thermal may be substituted when its efficiency compared to the thermal receiver is known for all frequencies. By using such an audiometer the average absolute sensitivity for approximately 100 ears which were considered to be normal was determined. The lower curve in

Fig. 1, labelled the threshold of audibility, shows the results of such measurements. The ordinates give the amplitude of the pressure variation in dynes per square centimeter that is just sufficient to cause an auditory sensation and the abscissæ give the frequency of vibration of the tone being perceived. Both are plotted on a logarithmic scale. The experimental difficulties made it impossible to make a very accurate determination for those parts of the curve shown by dotted lines. More work needs to be done on these portions of the curve. In the important speech range, namely, from 500 to 5,000 cycles, it requires approximately .001 of a dyne pressure variation in the air to cause an auditory sensation. This corresponds to a fractional change of about one-billionth in the atmospheric pressure, which shows the extreme sensitiveness of the hearing mechanism.

In order to obtain an idea of the intensity range used in hearing, an attempt was also made to obtain an upper limit for audible intensities. When the intensity of a tone is continually increased, a value is reached where the ear experiences a tickling sensation. Experiments show that the intensity for this sensation is approximately the same for various individuals and the results can be duplicated as accurately as those for the minumum intensity value. It was found that if this same intensity of sound is impressed against the finger, it excites the tactile nerves, In other words, the sensation of feeling for the ear is practically the same as for other parts of the body. When the intensity goes slightly above this feeling point, pain is experienced. Consequently, this intensity for the threshold of feeling was considered to be the maximum intensity that could be used in any practical way for hearing. The two points where these two curves intersect have interesting interpretations. At these two points, the ear both hears and feels the tone. At frequencies above the upper intersecting point, the ear feels the sound before hearing it, and in general would experience pain before exciting the sensation of hearing. Consequently, the intersection point may be considered as the upper limit in pitch which can be sensed. In a similar way, the lower intersection point represents the lowest pitch than can be sensed.

There has been considerable work [32-51] in the past to determine the upper frequency and lower frequency limits of audibility, but it would appear that without the criterion just mentioned, such limiting points apply only to the particular intensity used in the determination. Not enough attention has been paid to the intensity of the tones for such determinations. It is quite evident from this

figure that both the upper and lower limits of audibility which are found in any particular experimental investigation will very largely depend upon the intensity of the tones sounded. For example, if the intensity were along the .01 dyne line, the limits would be 200 and 12,600 cycles.

The area enclosed between the maximum and minimum audibility curves has been called the auditory-sensation area and each point in it represents a pure tone. The question then arises: How many such pure tones can be sensed by the normal ear?

The answer to this question has been made possible by the recent work of Mr. V. O. Knudsen.[14] In this work Knudsen made determinations of the sensibility of the ear for small differences in pitch and intensity. In Fig. 2, the average results of his measurements for

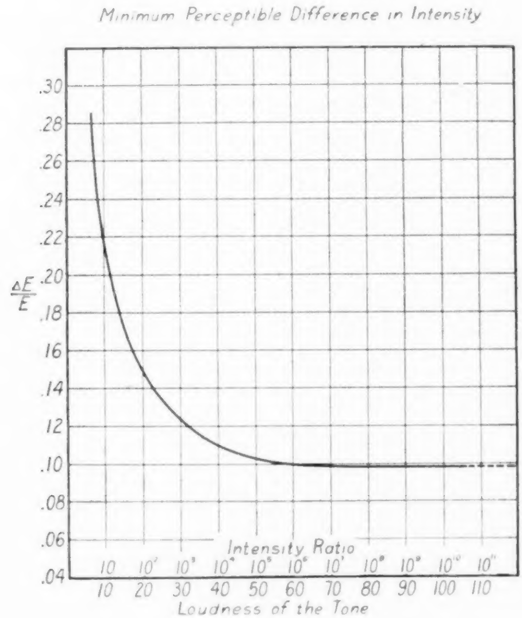


Fig. 2

changes in intensity are shown. Each ordinate gives the fractional change in the sound energy which is just perceptible, this fractional change being called the Fechner ratio. The abscissæ are equal to ten times the logarithm of the ratio of intensities, the zero corre-

sponding to the intensity at the threshold of audibility. For intensities greater than 10^4 times the threshold of audibility, the Fechner ratio has the constant value of approximately one-tenth. It was found that this ratio is approximately the same for all frequencies. In Fig. 3 is shown the results taken from Knudsen's article on the

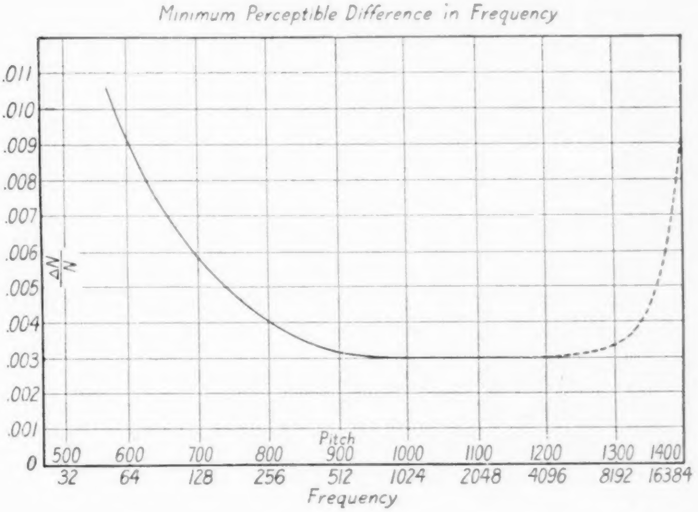


Fig. 3

pitch sensibility. The ordinates give the fractional change in the frequency which is just perceptible and the abscissæ give the frequency on a logarithmic scale. The meaning of the pitch scale at the bottom of this figure will be discussed later. For frequencies above 400 this fractional change is a constant equal to .003. This ratio probably becomes larger again for the very high frequencies. It was found that it varied with intensity in approximately the same way as that given for the energy ratio.

Using these values in connection with the auditory-sensation area, it is possible to calculate the number of pure tones which the ear can perceive as being different. For example, if, starting at the minimum audibility curve, ordinate increments are laid off along a constant pitch line, that are successively equal to the value of ΔE at the intensity position above the threshold, then the number of such increments between the upper and lower curves in Fig. 1 is equal to the number of pure tones of constant pitch that can be per-

ceived as being different in volume. If the minimum and maximum audibility curves were plotted on an energy scale, the increment length ΔE near the maximum audibility curve would be a million million times longer than its length in the minimum audibility curve, whereas when they are plotted on a logarithmic scale, this increment length remains approximately constant, changing by less than a factor 2 for 90 per cent. of the distance across the auditory-sensation area. The calculation shows (see Appendix A) that the number of such increments on the 100-cycle frequency line is 270, that is, 270 tones having a frequency of vibration of 1,000 cycles can be perceived as being different in loudness.

What has been said of the intensity scale applies equally well to the frequency scale. The calculation (see Appendix A) indicates that the number of tones that are perceivable as being different in pitch along the 10-dyne pressure line is approximately 1,300.

If an ordinate increment corresponding to ΔE and an abscissa increment corresponding to ΔN be drawn, a small rectangle will be formed which may be considered as forming the boundary lines for a single pure tone. All tones which lie in this area sound alike to the ear. The number of such small rectangles in the auditory-sensation area corresponds to the number of pure tones which can be perceived as being different. The calculation (see Appendix A) of this number indicates that there are approximately 300,000 such tones.

One might well ask the question: How many complex sounds which are different can be sensed by the ear? At first thought, one might say that this number is represented by all the possible combinations of pure tones. Of course, such a number would be entirely too large, for some of these would sound alike to the ear, since the louder tones would necessarily mask the feebler ones. It is evident, however, that the number of such complex sounds will be very much larger than the number of pure tones.

Scales of Loudness and Pitch

It is seen that the use of the logarithmic scale in Fig. 1 is much more convenient not only on account of the large range of values necessary to represent the auditory-sensation area, but also because of its scientific basis. Psychologists have recognized this since Weber and Fechner formulated the relation between the sensation and the stimulus. Although logarithmic units have been used by various authors in measuring the amount of sensation, the numerical values have been quite different. It seems inevitable that there will be a

greater cooperation in the future between men in the various branches of science working on this subject, so, in order to avoid misunderstanding, it would be very advantageous for all to use, as far as possible, the same units. With this in mind, I am taking the liberty of suggesting for discussion units for both loudness and pitch.

In the telephone business, the commodity being delivered to the customers is reproduced speech. One of the most important qualities of this speech is its loudness, so it is very reasonable to use a sensation scale to define the volume of the speech delivered. At the present time, an endeavor is being made to obtain an agreement of all the telephone companies, both in the United States and abroad, to adopt a standard logarithmic unit for defining the efficiency of telephone circuits and the electrical speech levels at various points along the transmission lines. The *chief interest* in changes in efficiency of transmission apparatus is their effects upon the loudness of the speech delivered by the receiver at the end of the telephone circuit. So it would be very advantageous to use this same logarithmic scale for measuring differences in loudness.

This scale is chosen so that the loudness difference is ten times the common logarithm of the intensity ratio. This means that if the intensity is multiplied by a factor 10, the loudness is increased by ten; if the intensity is multiplied by 100, the loudness is increased by 20; if the intensity is multiplied by 1,000, the loudness is increased by 30, etc. It was seen above that under the most favorable circumstances a change in loudness equal to 1/2 on this scale could just be detected. Knudsen's data indicate, however, that when a silent interval of only two seconds intervenes between the two tones being compared, a loudness change greater than unity on this scale is required before it is noticeable. So the smallest loudness change that is ordinarily appreciated is equivalent to one unit on this scale. It is also convenient because of the decimal relation between loudness change and intensity ratio. This relation is expressed by the formula:

$$\Delta L = L_1 - L_2 = 10 \log_{10} \frac{I_1}{I_2} \text{ or } \frac{I_1}{I_2} = 10^{\frac{\Delta L}{10}}$$

where L_1 and L_2 are the two loudness values corresponding to the intensities I_1 and I_2. Since intensities of sound are proportional to the square of pressure amplitudes this may also be written:

$$\Delta L = 20 \log \frac{p_1}{p_2}$$

The most convenient choice of the intensity or pressure used as a standard for comparison depends upon the problem under consider-

ation. In the sensation area chart of Fig. 1, the intensity line corresponding to one dyne was used as the zero level, that is, p_2 was chosen equal to 1 so that

$$\Delta L = 20 \log p$$

The choice of the base of logarithms for the pitch scale is dictated by the fact mentioned before, that the ear perceives octaves as being very similar sensations. Consequently the base 2 is the most logical choice for expressing pitch changes. If the logarithm of the frequency to the base 2 were used, perceptible changes in pitch would correspond to inconveniently small values of the logarithm. It is better to use the logarithm to the base $\sqrt[100]{2}$ which is 100 times as large. On this scale the smallest perceptible difference in pitch is approximately unity—somewhat more for frequencies greater than 100 cycles or somewhat less for lower frequencies, according to Knudsen's data. The scale on the charts is chosen so that the change in pitch is given by

$$\Delta P = 100 \log_2 N$$

where N is the frequency of vibration.

It is now evident why such pitch and loudness scales were used in Fig. 1. With these scales, the number of units in any area gives approximately the number of tones that can be ordinarily appreciated in that area. For example, there are approximately 2,000 distinguishable tones in each square, there being more near the centre and fewer near the boundary lines than this number.

Experiments have shown that pure tones of different frequencies which are an equal number of units above the threshold value sound equally loud. This statement may require modification when very loud tones are compared, but the data indicated that throughout the most practical range this was true. Consequently, the absolute loudness of any tone can be taken as the number of units above the threshold value.

Loudness of Complex Sounds

In the measurement of the loudness of complex tones, the situation is not so simple. It has been found that if two complex tones are judged equally loud at one intensity level and then each is magnified equal amounts in intensity, they then may or may not sound equally loud. The curves shown in Figs. 4, 5 and 6 will illustrate this. The first (Fig. 4) shows the comparisons at different intensity levels of two sounds whose pressure spectra are shown in the two figures at

the top. The x-axis gives units above threshold for sound A and the y-axis gives the units above threshold of the sound B when the two sound equally loud. In this case, the spectra are somewhat similar

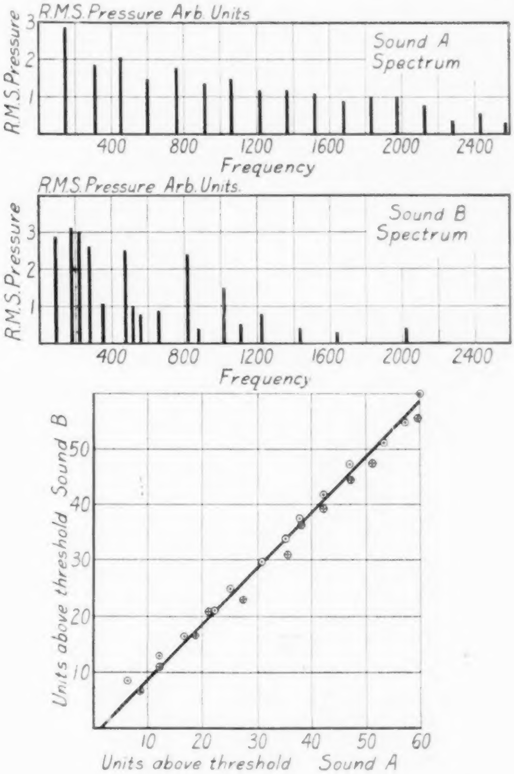


Fig. 4

and we have a straight line of slope 45° passing nearly through the origin. The two sounds are thus of practically equal loudness when they are the same number of units above threshold. In Fig. 5 we have similar data for two sounds which have quite different spectra as is indicated by the two charts at the top. The curve for C means that it was a practically continuous spectrum. It was produced by a device for making the "swishing" type of noises which are usually so prominent in office rooms. The curve representing the relation is not straight, since for values of intensity near the threshold, the

loudness increases faster for the C sound for increments in the intensity than for the A sound. For example, it is seen that when the sound C is 30 units above the threshold, the sound A is 45 units

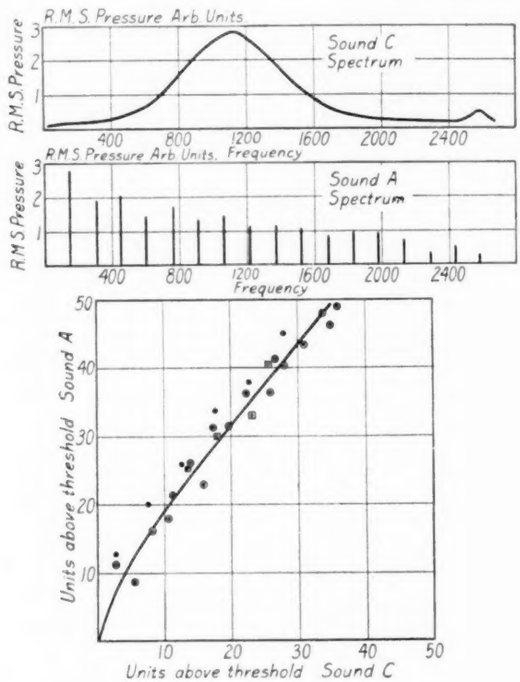


Fig. 5

above the threshold when the two sound equally loud. In Fig. 6 a comparison is given between the loudness of a pure tone of 700 cycles and a complex sound designated by A in the last figure. In this case again the relation is expressed by a curve. The technique of making such loudness measurements is rather difficult and requires a large number of observations before the values are reliable. A paper on this subject which will soon be published will give a detailed account of this work on loudness.

Enough data have been given to show that in order to give loudness a definite meaning for complex sounds, a more precise definition is necessary. It has been found convenient to define the loudness of any complex or pure tone in terms of the loudness of a sound standard.

This standard is a pure tone having a vibration frequency of 700 cycles per second. Its absolute loudness is defined as the change in loudness measured on the scale defined above, from the loudness value corresponding to the threshold pressure for normal ears which for 700 cycles is exactly 0.001 dyne. This frequency was arbitrarily chosen as a standard for measuring loudness because of this particular value of its threshold pressure, and because it is close to the frequency

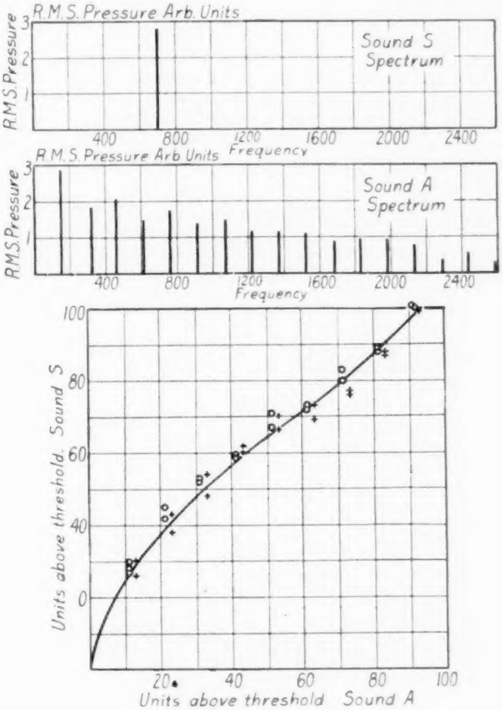


Fig. 6

at which the loudest tones used in conversational speech occur. By this definition, the loudness of a tone of frequency 700, for which p is the pressure variation, expressed as a root mean square value,

$$L = 60 + 20 \log p$$

and the loudness of any other sound, pure or complex, is defined as being equal to that of a tone of frequency 700, seeming equally loud. Such a definition implies that experimental measurements can be

made to determine when any complex sound is equally loud to a 700-cycle tone. Such measurements can be made although the observational error is rather large and the judgment of various individuals is sometimes quite different, which means only that loudness as measured by various individuals is different. For use in engineering work, however, the average of a large number of individuals can be taken and this loudness will have a definite determinable value. For example in Fig. 6, the loudness of the A sound when it is 60 units above the threshold is 72, since it sounds as loud as a 700-cycle tone which is 72 units above its threshold. The loudness of complex sounds usually increases faster with increases in intensity than that of pure tones. This would be expected since the threshold is determined principally by the loudest frequency in the complex sound and as the intensity is increased the other frequencies begin to add to the total loudness.

Since pure tones of different pitches which are the same number of units above the threshold sound equally loud their loudness L can be represented by the formula

$$L = L_0 + 20 \log p$$

where p is the root mean square value of the pressure amplitude produced in the ear by the tone and L_0 is the number of units from the 1-dyne line to the minimum audibility curve. The values of L_0 can be read directly from the chart in Fig. 1.

Measurement of Degree of Deafness

The choice of the loudness and pitch units used above leads to a *rational definition of the degree of deafness.*

The number of possible pure tones that can be sensed by a deaf person is considerably smaller than that mentioned above obtained from the normal auditory-sensation area. A logical way of defining the amount of hearing is: *To give the per cent. of the total number of distinguishable pure tones audible to a person with normal hearing, that can be sensed by the deaf person.*

Some definition of this sort will be very helpful in clearing up the confusion that now exists in court cases involving the degree of deafness. It is well known that there are a number of laws which prevent people who have more than a defined amount of deafness from doing certain classes of work. For example, one cannot operate an automobile if he has a certain per cent. of deafness. At the present time, there is a large variation between the standards set up by the various doctors in different parts of the country.

From the discussion above it was seen that the number of tones corresponding to any region was approximately proportional to the area of that region when the logarithmic units were used. Consequently the per cent.* of hearing can be taken as the fractional part

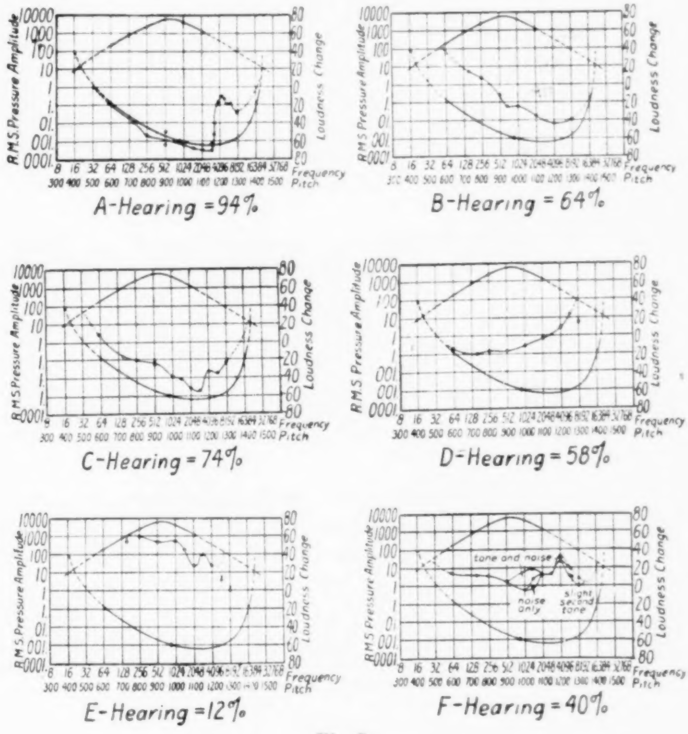


Fig. 7

Audiograms for Typical Cases of Deafness

of the normal auditory-sensation area in which tones can be properly sensed. The per cent. of deafness is, of course, 100 minus the per cent. of hearing.

To emphasize the meaning of this definition, some audiograms, that is minimum audible intensity curves, for some typical cases of deafness will be given. These are shown in Fig. 7. The first chart

* This assumes that the Fechner ratio for pitch and loudness is approximately the same for one having abnormal as for one having normal hearing.

shows a common type of deafness in which the sensitivity to the high frequencies suddenly decreases, as is indicated by the rise in the minimum audible intensity curve when the frequency exceeds 3,000 cycles per second. The sensation area for this person is 94 per cent. of that for the average. Consequently, his per cent. of hearing is 94 per cent. It is also convenient to speak of the per cent. of hearing for each pitch. It is evident that the logical definition for this is the ratio of the widths of the sensation area for the person tested and normal person, measured along the ordinate drawn at the frequency in question.[56] For example, in this audiogram the person had more than 100 per cent. hearing for most of the pitch range. At 4,000 cycles, however, the per cent. hearing was only 60 per cent. This means that for this pitch, the person when compared with one having normal hearing could sense only 60 per cent. as many gradations in tonal volume before reaching the threshold of feeling.

The second chart corresponds to a type of deafness that is not so common. It shows relatively large losses at the lower frequencies. The per cent. hearing in this case is seen to be 64 per cent.

The third type is very common and corresponds to a general lowering of the frequencies throughout the entire pitch range. In these first three cases, the deaf persons could carry on a conversation without any difficulty whatever. In the last two of these, difficulty was experienced in understanding a speaker at any considerable distance. In the first case, the person could not hear the steam issuing from a jet or any other high hissing sound. However, he could hear and understand speech practically as well as anyone with normal hearing.

The fourth case shows a falling off at the high frequencies, but this loss in hearing proceeds gradually as the pitch increases rather than abruptly as in the first case. As indicated in the figure the per cent. of hearing is 58 per cent.

The fifth case is one of extreme deafness and is typical of such cases. The per cent. of hearing is only 12 per cent. The last case shows not only the minimum audibility curve, but the quality of the sensation perceived. As indicated on the chart, in certain regions noises are heard when the stimulus is a pure tone. When computing the per cent. of hearing in such cases, it seems reasonable to take only the area where sensation of good quality is perceived. In some cases, this poor quality extends through practically the whole area and although the person hears sounds, he is unable to properly interpret them. Consequently, from a practical point of view, his per cent. of hearing is very low. For such cases, deaf sets or other aids to the hearing do not give any satisfactory help.

Masking of One Pure Tone by Another

We are now in a position to discuss another set of facts concerning the perception of tones, namely, the ability of the ear to perceive certain sounds in the presence of other sounds. Such data for pure tones have been obtained in our laboratories and will soon be published in some detail. The apparatus used consisted simply of two vacuum tube oscillators generating the two tones used and two attenuators which made it possible to introduce the tones into a single receiver with any desired intensities. In other words, it consists of two audiometers with a common receiver for generating the two tones. The curves shown in Fig. 8 give the general character of the results of this work.

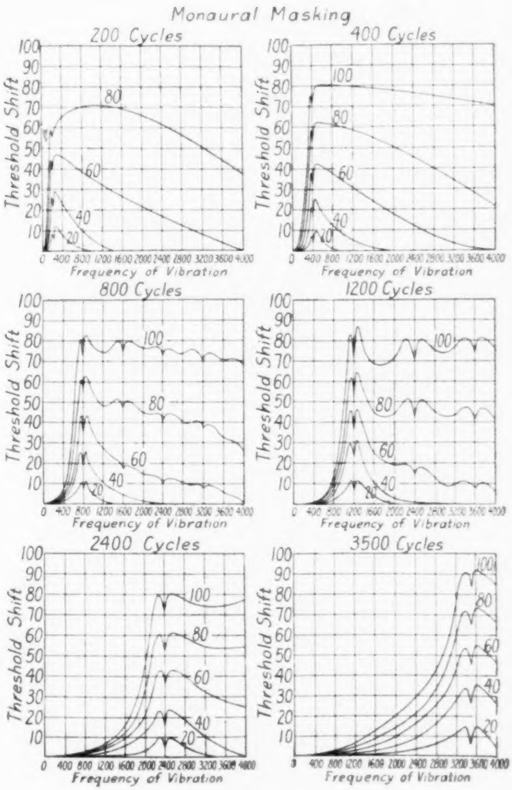


Fig. 8

The ordinates show the amounts in loudness units that the threshold value of a tone of any frequency called the "masked tone" is shifted due to the presence of another tone called the "masking tone." The frequency of the masking tone is given at the top of each set of curves.

The experimental procedure was as follows: The threshold values for the two tones were first determined. The intensity of the masking tone (the frequency of which is given above each graph) was then increased beyond its threshold value by the number of units indicated just above the curve. The masked tone was then increased in intensity until its presence was just perceived. The amount of this latter increase, measured on the loudness scale is called the *threshold shift* and is plotted as ordinate in Figs. 8, 9 and 10. The frequencies of the masked tones are given by the abscissæ.

For example, in the fourth chart, the masking effects of the tone having a frequency of 1,200 cycles are shown. It is seen that the greatest masking effect is near 1,200 cycles, which is the frequency of the masking tone. A tone of 1,250 cycles must be raised to 46 units above the threshold to be perceived in the presence of a 1,200-cycle tone which is 60 units above its threshold, or it must be raised to within 14 units of the masking tone before it is perceived. This corresponds to an intensity ratio between the tones of only 25. A tone of 3,000 cycles, however, can be perceived in the presence of a 1,200-cycle tone which is 60 units loud when it is only 8 units above its threshold. This means that the intensity ratio between these two tones, under such circumstances, corresponds to 52 units or to a ratio of approximately 160,000 in intensity. However, as the loudness of the masking tone is increased, all of the high tones must be increased to fairly large values before they can be heard. For example, the high frequencies must be raised 75 units above the threshold to be heard in the presence of a 1,200-cycle tone having a loudness of 100 units. But even for such large intensities for the masking tone, those frequencies below 300 are perceived by raising their loudness only slightly above the threshold value. It should be noticed that in all cases, those tones having frequencies near the masking frequency, whether they are higher or lower, are easily masked.

It is thus seen that Mayer's conclusion, that a low pitch sound completely obliterates higher pitched tones of considerable intensity and that higher pitched frequencies will never obliterate lower pitched tones, is true only under certain circumstances. A low tone will not obliterate to any degree a high tone far removed in frequency, except when the former is raised to very high intensities. Also a tone of higher frequency can easily obliterate a tone of lower fre-

quency if the frequencies of the two tones are near together. When the two tones are very close together in pitch the presence of the masked tone is perceived by the beats it produces. This accounts for the sharp drop in the curves at these frequencies. A similar thing happens for those regions corresponding to harmonics of the masking frequency. In the charts for the 200- and 400-cycle masking tones these drops are not shown inasmuch as they were small, but in an accurate picture they should be shown.

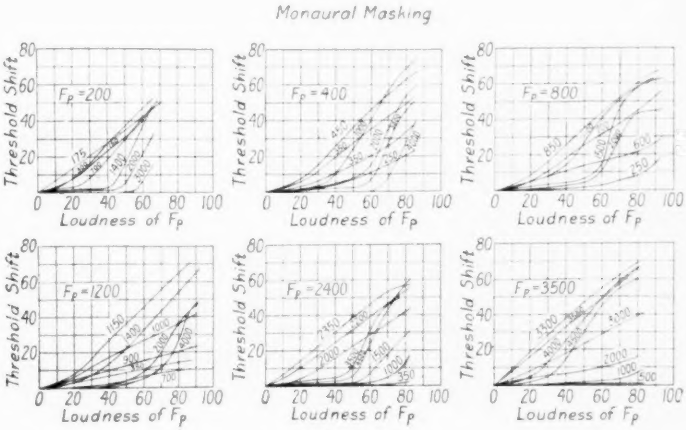


Fig. 9

In Fig. 9, these results are shown plotted in a different way. The abscissæ represent the loudness of the primary tones whose frequency is indicated at the top of each of the charts. The amounts that the threshold is shifted are plotted as ordinates as in the previous figure. For example, in Chart 1, the results are shown for a masking tone of 200 cycles. The curve marked 3,000 indicates the masking effect of a 200-cycle upon a 3,000-cycle tone. It is seen that the loudness of the low pitched tone can be raised to 55 units before it has any interfering effect upon the high pitched tone. For louder values than this it has a very marked effect. It will be noticed that in nearly all of the charts the curves for different frequencies intersect. This leads to some rather interesting conclusions, regarding the perception of a complex tone. For example, consider the curves for a masking tone having a frequency of 400 cycles. Assume we have a complex tone having three frequencies of 400, 300 and 200 cycles with relative loudness values of 50, 10 and 10, respectively. The ear will hear only

the 400-cycle tone and the 200-cycle tone as is evident from the curves. It would be necessary to raise the 300-cycle tone above 16 units for it to be heard in the presence of 400 cycles of loudness 50. However, if the sound is magnified without distortion 30 loudness units, so that these three frequencies have loudness values of 80, 40

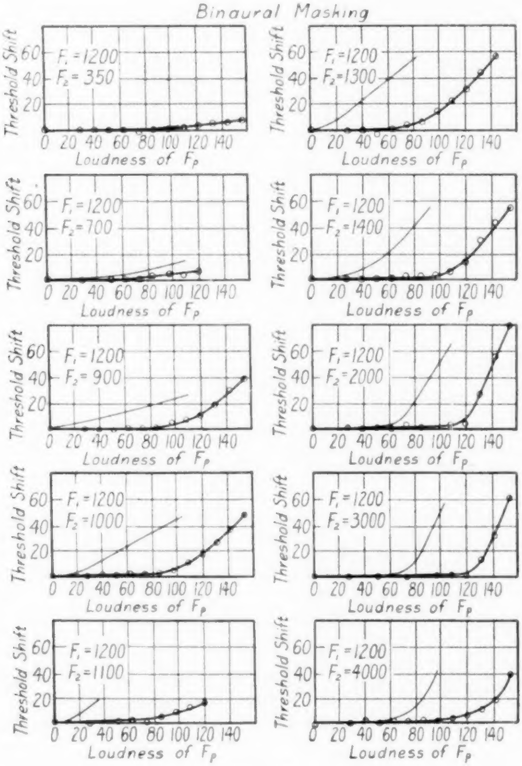


Fig. 10

and 40, respectively, then the 400-cycle tone and 300-cycle tone only will be heard. Under such conditions, the 300-cycle tone could be attenuated approximately 15 units before it would disappear. This means that the sensation produced by a complex sound is different in character as well as intensity when the sound is increased or decreased in intensity without distortion. In general, as the tone becomes more intense the low tones become more prominent because the high

tones are masked. It is a common experience of one working with complex sounds to have the low frequencies always gain in prominence as the sound is amplified.

The question naturally arises, Does the same interfering effect exist when the two tones are introduced into opposite ears instead of both being introduced into the same ear? The answer is No. Curves showing the results in such tests are shown in Fig. 10. For comparison the results for the case when in tones are both in the same ear are given by the light lines. Take the case of 1,200 and 1,300 cycles. It is rather remarkable that a tone in one ear can be raised to 60 units, that is, increased in intensity one million times, before the threshold value for the tone in the other ear is noticeably affected. If the 1,300-cycle tone were introduced into the same ear as the 1,200-cycle tone, its loudness would need to be shifted 40 units, corresponding to a 10,000-fold magnification in intensity above its threshold intensity in the free ear before it can be heard. It is seen that if one set of curves is shifted about 50 units it will coincide with the second set. This strongly suggests that the interference in this case is due to the loud tone being transmitted by bone conduction through the head with sufficient energy to cause masking. The vibration is probably picked up by the base of the incus and transmitted from there to the cochlea in the usual way. There is other evidence * which I shall not have space here to discuss, which indicates that the effective attenuation from one ear to the other is approximately 50 units.

Theories of Audition

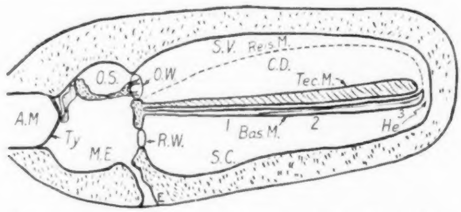
With these facts in mind, we are now ready to discuss the theory of hearing which will best account for them. I will refer briefly to just a few of the principal theories of hearing which have been proposed. The sketch shown in Fig. 11 gives a diagrammatic picture of the internal ear. In the Helmholtz theory, as first formulated, it is stated that the organ of Corti located between the basilar membrane and the tectorial membrane act like a set of resonators which are sharply tuned. Each tone stimulates a single organ depending upon its pitch. Later this theory was somewhat modified as it was thought that the resonant property might reside in one of the membranes in the cochlea.

* See paper by Wegel and Lane soon to be published in the *Physical Review* entitled "The Auditory Masking of One Pure Tone by Another and its Relation to the Dynamics of the Inner Ear."

In the "telephone" theory, as expounded by Voltalini, Rutherford, Waller and others, it is assumed that the basilar membrane vibrates as a whole like the diaphragm of a telephone receiver, and consequently responds to all frequencies with varying degress of amplitude. The discrimination of pitch takes place in the brain.

Meyers in his theory states that various lengths of the basilar membrane are set in motion depending upon the intensity of the stimulating tone. As in the previous theory, the pitch discrimination is accomplished in some way in the brain.

In the "non-resonant" theory of Emile ter Kuile it is assumed that the sound disturbance penetrates different distances into the



Diagrammatic representation of auditory function

A.M.	Auditory meatus	O. W.	Oval window
Bas. M.	Bas. mem. including organ of Corti	Reis. M.	Reissner's mem.
C. D.	Cochlear duct	R. W.	Rd. window
E.	Eustachian tube	S. C.	Scala cochlea
He.	Helicotrema	S. V.	Scala vestibuli
M. E.	Middle ear	Tec. M.	Tectorial membrane
O. S.	Ossicles (malleus, incus, stapes)	Ty.	Tympanic membrane

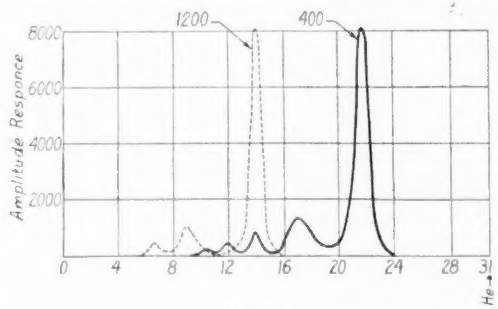


Fig. 11

cochlea depending upon the frequency of the stimulating tone. The further along the membrane the disturbance reaches, the lower will be the pitch sensation. A low pitch tone then stimulates all of the nerve fibres that would be stimulated by tones of higher pitch plus some additional nerve fibres.

The theory of maximum amplitudes was first put into definite form by Gray in 1899.[69] It assumes that the position of maximum amplitude of the basilar membrane varies with the pitch of the stimulating tone. Although a considerable portion of the membrane vibrates when stimulated by a pure tone, the ear judges the pitch by the position of maximum response of the basilar membrane. Roaf has shown that some action of this sort must take place due to the dynamical constants involved.[61] It is an amplification of this theory that I desire to propose as the one which most satisfactorily accounts for the facts.

When a sound wave impinges upon an ear-drum, its vibrational motion is communicated through the middle ear (Fig. 11) by means of the chain of small ossicles (malleus, incus and stapes) to the oval window. Here the vibration is communicated to the fluid contained in the cochlea. If the pitch of the tone is low, say below 20 vibrations, the fluid is moved bodily back and forth around the basilar membrane through the helicotrema, the motion of the membrane at the round window and the oval window being just opposite in phase, the former moving inward while the latter moves outward. For very high frequencies, the mass reactions of the ossicles and the fluid are so great that very little energy can be transmitted to the cochlea. For example, when the elastic forces are negligible it requires a force 10,000 times as large as produce the same amplitude of vibration at 10,000 cycles as that required at 100 cycles. For intermediate frequencies the mass reactions, the elastic restoring forces and the frictional resistances which are brought into play are such that the wave is transmitted through the basilar membrane causing the nerves to be excited.

It is thus seen that the upper and lower limits of audibility are easily explained. When the forces upon the drum of the ear or walls of the ear canal are large enough to excite the sensation of feeling and the pitch of the tone is either too low or too high to cause any perceptible vibration of the basilar membrane, we are beyond the lower or upper limit of audibility respectively. At frequencies between these limits, the vibrational energy is first communicated to the fluid in the scala vestibuli and then transmitted through the two membranes into the fluid of the scala cochlea. As the basilar membrane transmits the sound wave it takes up a vibration amplitude which stimulates the nerve fibres located in it. The entire membrane vibrates for every incident tone, but for each frequency there is a corresponding spot on the membrane where the amplitude of

the vibration is greater than anywhere else. Our postulate is that only those nerves are stimulated which are at the particular parts of the membrane vibrating with more than a certain critical amplitude; and that we judge the pitch from the part of the membrane where the nerves are stimulated. According to this conception, the variation with frequency of the minimum audible intensity is due principally to the variation with frequency of the transmission efficiency of the mechanical system between the auditory meatus and the basilar membrane. Pure tones of equal loudness correspond either to equal amplitudes or to equal velocities of vibration of the basilar membrane or to some function of the two. Whatever is assumed, the dependence of the minimum audible intensity upon frequency for the ear can be explained entirely by the vibrational characteristics of the ear mechanism. For the sake of clearness it will be assumed that equal amplitudes of vibration of the basilar membrane correspond to equal sensations. For loud pure tones, there are several regions of maximum amplitude on the membrane, corresponding to the tone and to the harmonic introduced by the non-linear response of the middle ear, the latter maxima increasing very rapidly as the stimulation increases.

It is a strange thing that the phenomenon of the masking of tones which, as stated in the beginning, has been considered by some to be so fatal to any resonator theory, is the very thing that has furnished experimental data which makes it possible to calculate the vibration characteristics of the inner ear. Such a calculation must be based upon assumptions which will be uncertain, but will seem reasonable. It is not my purpose to discuss those here, but I shall give only the final result of such a calculation made by Mr. Wegel and Mr. Lane of our laboratories. At the bottom of Fig. 11, the two curves show the amplitude of vibration of different portions of the basilar membrane for the two frequencies 400 and 1,200 cycles. For purposes here these curves may be considered to be simply illustrative. This membrane has a length of 31 mm. and a width of .2 mm. at the base and .36 mm. at the helicotrema end. The x-axis in this figure gives the distance in millimeters from the oval window and the y-axis gives the amplitudes of vibration in terms of the amplitude corresponding to the threshold of audibility. The loudness of the stimulating tones in both cases is 80 units. It will be seen that the maximum response for the high frequencies is near the base of the cochlea, while that for the low frequencies is near the helicoptrema. It will be noticed that the amplitude of the membrane has several maxima corresponding to the subjective harmonics.

With this picture in mind, it is clear why the perception of one tone is interfered with by the presence of a second tone when their frequencies are close together, since the nerves necessary to perceive the first tone are already stimulated by the second tone. Also when their frequencies are widely separated, entirely different sets of nerves carry the impulses to the brain, and consequently there is no interference between the tones except that which occurs in the brain. Although this brain interference may not be entirely negligible, especially for very loud sounds, it is certainly very much smaller than that existing in the ear for tones close together in pitch.

It is also seen that the reason why the low tones mask the high tones very much more easily that the reverse is due to the harmonics introduced by the transmission mechanism of the ear. Inasmuch as these harmonics are due to the second order modulations, they are proportional to the square of the amplitude and, therefore, become much more prominent for the large amplitudes. When two tones are introduced, summation and difference tones as well as the harmonics will necessarily be present (see Appendix B). With the proper apparatus for generating continuously sounding tones, these subjective tones are easily heard. Their frequency can be quite accurately located by introducing from an external source a frequency which can be varied until it produces beats with the subjective tone.

Messrs. Wegel and Lane who are working in this field have observed modulation frequencies created in the ear as high as the fourth order. They will soon publish* an account of this work on the vibrational characteristics of the basilar membrane. It is seen that the quality as well as the intensity of the sensation produced by a pure tone should change as the intensity of stimulus is increased due to the increasing prominence of the harmonics. This is in accordance with one's experience while listening to pure tones of varying intensity. The non-linear character of the hearing mechanism is also sufficient to account for the falling off in the ability of one to interpret speech when it becomes louder than about 75 units. The introduction of the summation and difference tones and the harmonics makes the interpretation by the brain more difficult. Its action in this respect is very similar to the carbon transmitter used in commercial telephone work or to an overloaded vacuum tub. This characteristic of the ear also explains why we should expect departures from non-linearity when making loudness balances for complex tones. It also suggests that a similar thing might be ex-

* Wegel and Lane, see paper already cited.

pected when comparing the loudness of pure tones if the balances are made at very high intensities. No such balances have yet been made.

What happens to the ear when one becomes deaf? This question, of course, is one for the medical profession to answer, but let us take one or two simple cases and see if they fit into this theory. First assume that the nerve endings are diseased for a short distance away from the base of the cochlea so that they send no impulses to the brain. Under certain assumptions the kind of an audiogram one should obtain can be calculated from the vibrational characteristics determined as mentioned above. Such a calculation shows that an audiogram similar to that shown in Fig. 4, which has a rapid falling off in sensitiveness, can be accounted for, both quantitatively as well as qualitatively. On a pure resonant theory corresponding to that first proposed by Helmholtz, a tone island would exist corresponding to the affected region for such a case. Although we have tested a large number of cases, no such islands have ever been found. When the intensity of the tone is raised sufficiently to bring the amplitude of the area containing the healthy nerve cells which are adjacent to the diseased portion to a value above that corresponding to the threshold, the tone will then be perceived.

Again assume that due to some pathological condition, the tissue around the oval window where the stapes join the cochlea has become hardened. Its elasticity will then be greatly increased so that vibrational energy at low frequencies will be greatly discriminated against. For such a case, an audiogram similar to that shown in Fig. 7-B would be obtained.

A number of things can cause a general lowering of the ear sensitivity, such as wax in the ear canal, affections of the ear-drum, fixation of any of the ossicles, thickening of the basilar membrane, affections of the nerve endings or loss in nervous energy being supplied to the membrane, etc. However, one would expect that each type of trouble would discriminate, at least to some extent, against certain frequency regions so as to produce some characteristic in the audiogram. Ear specialists are beginning to realize the possibility of obtaining considerable aid in the diagnosis of abnormal hearing from such accurate audiograms.

There are a large number of facts obtained from medical research which necessarily have a bearing upon the theory of hearing, but as far as I know none of them is contrary to the theory of hearing given above. It was seen that there are approximately 300,000 tone units in the auditory-sensation area. According to the anatomists, there are

only 4,000 nerve cells in the basilar membrane with four or five fibre hairs for each cell. Assuming that each hair fibre acts as a unit there are still insufficient units for each perceivable tone and according to the theory given above, a large number of these units must act at one time. Consequently the ear must be able to interpret differences in the intensity of excitation of each nerve cell as well as determine the position of each nerve cell excited.

Most modern neurologists believe in the "none or all" excitation theory of nerve impulses.[59-60] They also claim that nerve impulses can never be much more rapid than about 50 per second and cannot therefore follow frequencies as high as those found in sound waves. The second statement only emphasizes the necessity of assuming that the intensity position as well as place position is necessary to

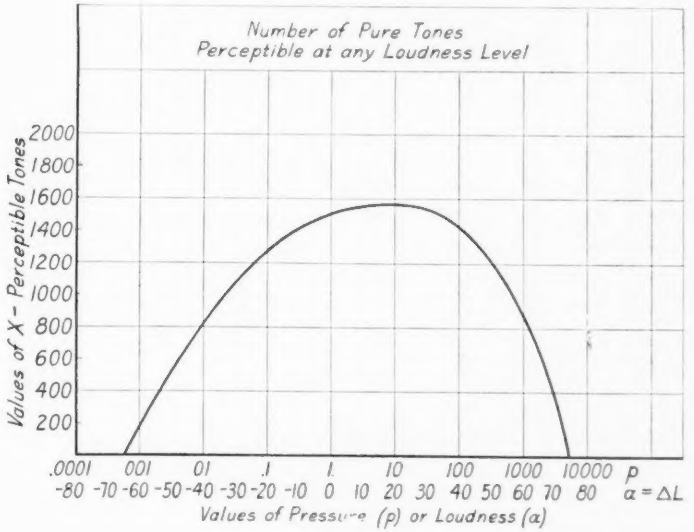


Fig. 12

account for the differentiation of pure tones. The first statement is not necessarily in conflict with such an idea since anatomists are not agreed upon the number of nerve fibres radiating from each nerve cell. Since each nerve fibre can serve to give a unit nerve impulse, the intensity of stimulation sent from a single nerve cell can increase with stimulation depending upon the number of nerve fibres brought into action. The intensity of the sensation produced

is then directly related to the total number of nerve fibres giving off impulses. It seems to me that the spacial and intensity configurations which are possible, according to this theory, are sufficient for an educated brain to interpret all the complex sounds which are common to our experience.

In conclusion then, it is seen that the pitch of pure tones is determined by the position of maximum response of the basilar membrane, the high tones stimulating regions near the base and the low tones regions near the apex of the cochlea.

A person can sense two mixed tones as being distinctly two tones while he cannot sense two mixed colors, since in the ear mechanism there is a spacial frequency selectivity while in the eye mechanism there is no such selectivity.

The limiting frequencies which can be perceived are due entirely to the dynamical constants of the inner ear as is also the dependence of minimum audible intensity on frequency.

The so-called subjective harmonics, summation and difference tones are probably due to the non-linear transmission characteristics of the middle and inner ear.

These subjective harmonics account for the greater masking effect of low tones on high tones than high tones on low tones. Due to this non-linear characteristic, the quality as well as the intensity of the sensation produced, especially by complex tones, change as the intensity of the stimulus increases.

The facts obtained from audiograms of abnormal hearing are consistent with the theory of hearing which has been outlined.

Although this theory of hearing involves the principle of resonance, it is very different from the Helmholtz theory as usually understood. In the latter it as assumed that there are four or five thousand small resonators in the ear, each responding only to a single tone; while in the former it is assumed that a single vibrating membrane which vibrates for every impressed sound is sufficient to differentiate the various recognizable sounds by its various configurations of vibration form.

A loudness scale has been chosen such that the loudness change is equal to ten times the common logarithm of the intensity ratio. A pitch scale has been chosen such that the pitch change is equal to 100 times the logarithm to the base two of the frequency ratio. The loudness of complex or simple tones is measured in terms of the number of loudness units a tone of 700 cycles must be raised above its average threshold value before it sounds equally loud to the sound measured.

The degree of deafness is measured by the fractional part of the normal area of audition in which the sensation is either lacking or false.

Appendix A

The calculations of the number of pure tones perceivable as being different in pitch at a given intensity or being different in loudness at a given pitch involves a line integral. The calculation of the number of pure tones perceivable as being different either in loudness or pitch involves a surface integral.

Let the coordinates used in Fig. 1 corresponding to ΔL and ΔP be designated α and β, respectively. Then the relations shown in Figs. 2 and 3 can be expressed by the equations

$$\frac{\Delta E}{E} = f(\alpha - \alpha_0) \text{ and} \tag{1}$$

$$\frac{\Delta N}{N} = \varphi(\beta) \tag{2}$$

where α_0 is the value of α along the normal minimum audibility curve shown in Fig. 1. Knudsen's data indicated that the curve shown in Fig. 3 held only for values of $\alpha - \alpha_0$ corresponding to the flat part of the curve in Fig. 2. For lower intensities the pitch discrimination fell off in about the same way as that shown for the intensity discrimination. To represent this mathematically, $\varphi(\beta)$ can be multiplied by a factor which is unity for the loud tones and which increases similarly to $f(\alpha - \alpha_0)$ for the weaker tones. Such a factor is $10\,f(\alpha - \alpha_0)$ since $f(\alpha - \alpha_0)$ is approximately $\frac{1}{10}$ for the louder tones. So the corrected formula for $\frac{\Delta N}{N}$ is

$$\frac{\Delta N}{N} = 10\varphi(\beta) \cdot f(\alpha - \alpha_0). \tag{3}$$

Let dx be the number of perceivable tones of constant intensity corresponding to α in the pitch region between β and $\beta + d\beta$ and let dy be the number of perceivable tones of constant pitch corresponding to β in the region between α and $\alpha + d\alpha$. Then

$$dx = \frac{dN}{\Delta N} \tag{4}$$

$$dy = \frac{dE}{\Delta E}. \tag{5}$$

But the values of β and α are given by

$$\beta = 100 \log_2 N \tag{6}$$

$$\alpha = 10(\log_{10} E - \log_{10} E_1) \tag{7}$$

where E_1 is the value of intensity corresponding to a pressure amplitude of 1 dyne.

Substituting values of dN and dE in terms of α and β we have

$$dx = \frac{1}{100}\frac{N}{\Delta N}\log_e 2\, d\beta = \frac{\log_e 2}{1000}\frac{d\beta}{\varphi(\beta)\cdot f(\alpha - \alpha_0)} \tag{4'}$$

$$dy = \frac{1}{10}\frac{E}{\Delta E}\log_e 10\, d\alpha = \frac{\log_e 10}{10}\frac{d\alpha}{f(\alpha-\alpha_0)}. \tag{5'}$$

The number of tones of constant intensity which are perceivable as different in pitch is then

$$x = \frac{\log_e 2}{1000}\int_{\beta_1}^{\beta_2}\frac{d\beta}{\varphi(\beta)\cdot f(\alpha - \alpha_0)}$$

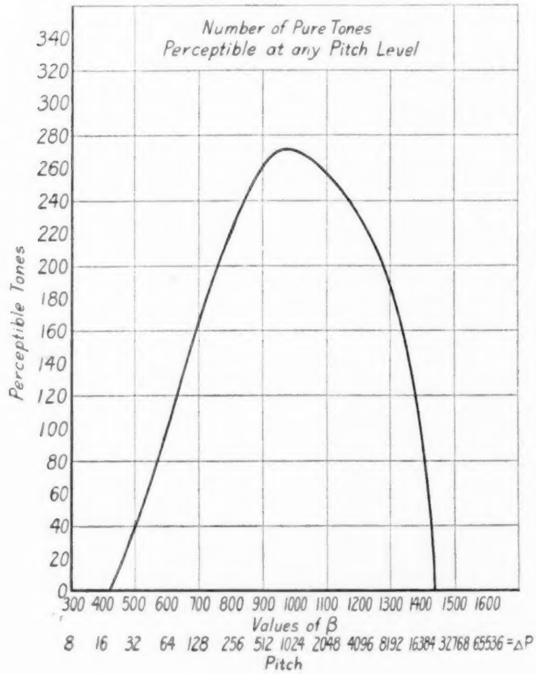


Fig. 13

where β_1 and β_2 are the points where the particular intensity line cuts the boundary lines of the auditory-sensation area. For example, the limits for the line corresponding to 1-dyne pressure ampli-

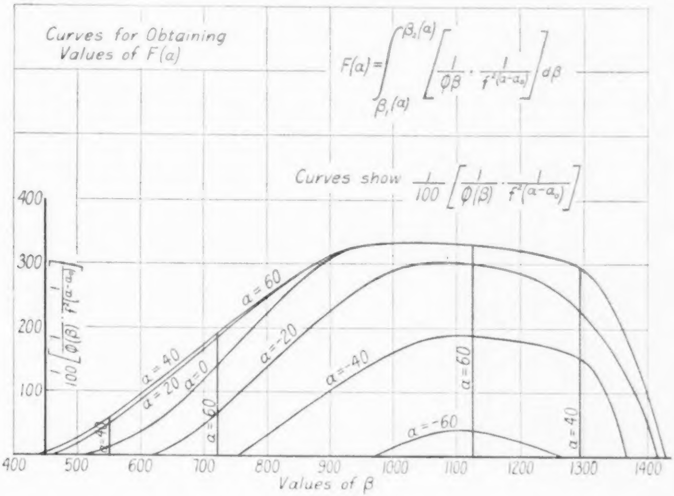


Fig. 14

tude are 500 and 1420. Similarly the number of tones of constant pitch which are perceivable as being different in loudness is given by

$$y=\frac{\log_2 10}{10}\int_{\alpha_1}^{\alpha_2}\frac{d\alpha}{f(\alpha-\alpha_1)}$$

where α_1 and α_2 are determined by the intersection of the particular pitch line with the boundary lines of the auditory-sensation area.

The values of these integrals were computed graphically. Figs. 12 and 13 show the results of these calculations. It is seen that the maximum number of tones perceivable as different in loudness is in the frequency range 700 to 1,500 which is also the important speech range. The number in this range is approximately 270.

In the pressure range from 1 to 100 there are approxmiately 1,500 tones which can be perceived as being different in pitch.

The number of tones ΔT in a small area $d\beta\ d\alpha$ situated with one corner at the point (α, β) is given by $dx\ dy$ or

$$\Delta T = dx\ dy = \frac{\log_e 2 \log_e 10}{10{,}000}\ \frac{d\alpha\ d\beta}{\varphi(\beta) f^2(\alpha-\alpha_0)},$$

$$T = \frac{\log_e 2 \log_e 10}{10{,}000}\int\int\frac{d\beta\ d\alpha}{\varphi(\beta) f^2(\alpha-\alpha_0)}.$$

The function $\dfrac{1}{\varphi(\beta)\cdot f^2(\alpha-\alpha_0)}$ must be integrated throughout the auditory-sensation area. This was done by graphical methods as shown in Figs. 14 and 15 with the result that $T=324{,}000$.

Appendix B

Let the pressure variation of the air in front of the drum of the ear be designated by δp. Since the pressure of the air in the middle

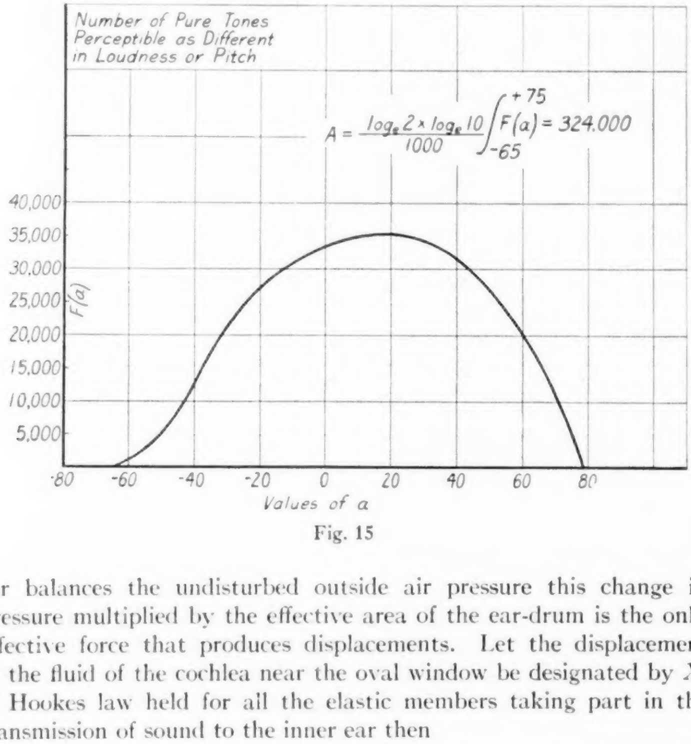


Fig. 15

ear balances the undisturbed outside air pressure this change in pressure multiplied by the effective area of the ear-drum is the only effective force that produces displacements. Let the displacement of the fluid of the cochlea near the oval window be designated by X. If Hookes law held for all the elastic members taking part in the transmission of sound to the inner ear then

$$X=k\delta p \tag{1}$$

where k is a constant.

It would be expected from the anatomy of the ear that Hookes law would start to break down even for small displacements. So in general the relation between the force δp and the displacement X can be represented by

$$X=f(\delta p)=a_0+a_1\delta p+a_2(\delta p)^2+a_3(\delta p)^3+\ \ldots . \tag{2}$$

where the coefficients $\alpha_0, \alpha_1, \alpha_2 \ldots$ belong to the expansion of the function into a power series. Now if δp is a sinusoidal variation then

$$\delta p = p_0 \cos \omega t \tag{3}$$

where $\frac{\omega}{2\pi}$ is the frequency of vibration. Substituting this value in (2), terms containing the cosine raised to integral powers are obtained. These can be expanded into multiple angle functions. For example, for the first four powers

$$\cos^2 \omega t = \tfrac{1}{2} \cos 2\,\omega t + \tfrac{1}{2}, \tag{4}$$

$$\cos^3 \omega t = \tfrac{1}{4} \cos 3\,\omega t + \tfrac{3}{4} \cos \omega t, \tag{5}$$

$$\cos^4 \omega t = \tfrac{1}{8} \cos 4\,\omega t + \tfrac{1}{2} \cos 2\,\omega t + \tfrac{3}{8}. \tag{6}$$

It is evident then that the displacement X will be represented by a formula

$$X = b_0 + b_1 \cos \omega t + b_2 \cos 2\,\omega t + b_3 \cos 3\,\omega t + \ldots .$$

In other words when a periodic force of only one frequency is impressed upon the ear-drum this same frequency and in addition all its harmonic frequencies are impressed upon the fluid of the inner ear.

If two pure tones are impressed upon the ear then δp is given by

$$\delta p = p_1 \cos \omega_1 t + p_2 \cos \omega_2 t.$$

If this value is substituted in equation (2), terms of the form $\cos^n \omega_1 t$ and $\cos^m \omega_2 t$ and $\cos^n \omega_1 t \cos^m \omega_2 t$ are obtained. The first two forms give rise to all the harmonics and the third form gives rise to the summation and difference tones. For example, the first four terms are

$$a_0 = a_0$$

$$a_1 \delta p = a_1 (p_1 \cos \omega_1 t + p_2 \cos \omega_2 t)$$

$$a_2 (\delta p)^2 = a_2 [\tfrac{1}{2} p_1^2 \cos 2\,\omega_1 t + \tfrac{1}{2} p_2^2 \cos 2\,\omega_2 t + p_1 p_2 \{\cos (\omega_1 - \omega_2) t + \cos (\omega_1 + \omega_2) t\} + \tfrac{1}{2} (p_1^2 + p_2^2)]$$

$$a_3 (\delta p)^3 = a_3 \Big[\left(\frac{3}{4} p_1^3 + \frac{3}{2} p_1 p_2^2 \right) \cos \omega_1 t + \frac{1}{4} p_1^3 \cos 3\,\omega_1 t + \left(\frac{3}{4} p_2^3 + \frac{3}{2} p_1^2 p_2 \right) \cos \omega_2 t + \frac{1}{4} p_2^3 \cos 3\,\omega_2 t + \frac{3}{4} p_1^2 p_2 \cos (\omega_2 t + 2\,\omega_1 t) + \frac{3}{4} p_1^2 p_2 \cos (\omega_2 t - 2\,\omega_1 t) + \frac{3}{4} p_1 p_2^2 \cos (\omega_1 t + 2\,\omega_2 t) + \frac{3}{4} p_1 p_2^2 \cos (\omega_1 t - 2\omega_2 t) \Big].$$

Therefore unless there is a linear relation between a force acting on the ear-drum and the displacement at the oval window, that is unless all the coefficients in equation (2) are zero except a_1, all the harmonics and the summation and difference tones will be impressed upon the fluid in the cochlea of the inner ear.

BIBLIOGRAPHY

Pitch Discrimination

[1] Preyer, "Grenzen d. Tonwahr.," Jena, 1876.
[2] Luft, *Phil. Studien.*, 4, p. 511, 1888.
[3] Meyer, *Ztschr. f. Psychol. u. Physiol.*, 16, p. 352, 1898.
[4] Schaefer and Guttman, *Ztschr. f. Psychol. u. Physiol.*, 32, p. 87, 1903.
[5] Stucher, *Ztschr. f. Psychol. u. Physiol.*, 42, part 2, p. 392, 1907.
[6] Seashore, *Psychol. Monogr.*, 13, p. 21, 1910.
[7] Vance, *Psychol. Monogr.*, 16, No. 3, p. 115, 1914.
[8] Smith, *Psychol. Monogr.*, 16, No. 3, p. 65, 1914.
[9] Knudsen, *Phys. Rev.*, 21, No. 1, p. 84, Jan., 1923.

Intensity Discrimination

[10] Merkel, *Phil. Studien.*, 4, pp. 117–251, 1887.
[11] *Wien. Ann. d. Phys.*, 36, p. 834, 1889.
[12] Zwaardemaker, *Proc. Konink. Akad. Wetensch, Amsterdam*, 8, p. 421, 1905.
[13] Pillsbury, *Psychol. Monogr.*, 13, No. 1, p. 5, 1910.
[14] Knudsen, *Phys. Rev.*, 21, No. 1, p. 84, Jan., 1923.

Absolute Sensitivity

[15] Toepler and Boltzman, *Ann. d. Phys.*, 141, p. 321, 1870.
[16] Wead, *Am. Jour. Sci.*, (3), 26, p. 177, 1883.
[17] Rayleigh, *Proc. Roy. Soc.*, 26, p. 248, 1887; *Phil. Mag.*, [5], 38, pp. 295, 365, 1894; *Phil. Mag.*, [6], 14, p. 596, 1907.
[18] *Wien. Archiv. fur die gesamte Physiologie*, 97, pp. 1–57, 1903.
[19] Zwaardemaker and Quix, *Arch. f. Anat. u. Physiol. Abt.*, p. 321, 1903.
[20] Webster, *Festschrif. f. L. Boltzmann*, Leipzig, p. 866, 1904.
[21] Shaw, *Proc. Roy. Soc.*, Ser. A, 76, 360, 1905.
[22] Abraham, *Compt. Rend.*, 144, p. 1099, 1907.
[23] Koehler, *Ztschr. f. Psychol. u. Physiol.*, 54, (1), p. 241, 1910.
[24] Bernbaum, *Ann. d. Phys.*, 49, pp. 201–228, 1916.
[25] Kranz, *Phys. Rev.*, 17, No. 3, p. 384, 1921.
[26] Minton, *Phys. Rev.*, 19, No. 2, p. 80, 1922.
[27] Hewlett, *Phys. Rev.*, p. 52, Jan., 1922.
[28] Fletcher and Wegel, *Phys. Rev.*, 19, No. 6, p. 553, 1922.
[29] Lane, *Phys. Rev.*, 19, No. 5, May, 1922.
[30] Wegel, *Proc. Nat'l Acad.*, July 15, 1922.
[31] MacKenzie, *Phys. Rev.*, 20, No. 4, Oct., 1922.

Upper Limit of Audibility

[32] Savart, *Compt. Rend.*, 20, pp. 12–14, 1845.
[33] Rayleigh, *Proc. Roy. Inst. of Gr. Br.*, 15, p. 417, 1897.
[34] Koenig, *Handb. d. Physiol.*, 3, p. 112, 1880; *Ann. d. Phys.*, 69, pp. 626–721, 1899.

[35] Scripture and Smith, *Stud. from Yale Psychol. Lab.*, 2, p. 105, 1894.
[36] Stumpf and Meyer, *Ann. d. Phys.*, 61, p. 773, 1897.
[37] Schwendt, *Arch. f. d. ges. Physiol.*, 75, p. 346, 1899.
[38] Edelmann, *Ann. d. Phys.*, 2, p. 469, 1900.
[39] Meyer, *Jour. Physiol.*, 28, p. 417, 1902.
[40] *Wien. Arch. f. ges. Physiol.*, 97, p. 1, 1903.
[41] Bezold, *Funk. Pruef. d. Mensch. Gehoer.*, 2, p. 162, 1903.
[42] Schulze, *Ann. d. Phys.*, 24, p. 785, 1907.
[43] Stueker, *Sitz. Ber. d. Akad. d. Wiss. Wien.*, 116, 2a, p. 367, 1907.
[44] Wegel, *Proc. Nat'l Acad.*, July 15, 1922.

Lower Limit of Audibility

[45] Savart, *Ann. de Phys. et de Chem.*, 47.
[46] Helmholtz, "Sens. of Tone," Eng. Trans., p. 175, 1885.
[47] Preyer, "Ueber d. Grenze d. Tonwahr," Jena, p. 8, 1876.
[48] Benzold, *Ztschr. f. Psychol. u. Physiol.*, 13, p. 161, 1897.
[49] Schaefer, *Ztschr. f. Psychol. u. Physiol.*, 21, p. 161, 1899.
[50] Imai and Vance, *Psychol. Monogr.*, 16, p. 104, 1914.
[51] Wegel, *Proc. Nat'l Acad.*, July 15, 1922.

Miscellaneous References

[52] Mayer, A. M., *Phil. Mag.*, 11, p. 500, 1875, "Researches in Acoustics."
[53] Peterson, *Jos. Psychol. Rev.*, 23, No. 5, p. 333, 1916, "The Place of Stimulation in the Cochlea *vs.* Frequency as a Direct Determiner of Pitch."
[54] Arnold and Crandall, *Phys. Rev.*, 10, No. 1, July, 1917, "The Thermophone as a Precision Source of Sound."
[55] Wente, *Phys. Rev.*, 19, No. 5, p. 498, 1922, "The Sensitivity and Precision of the Electrostatic Transmitter for Measuring Sound Intensities."
[56] Fletcher and Wegel, *Phys. Rev.*, 19, No. 6, June, 1922, "The Frequency Sensitivity of Normal Ears."
[57] Fowler and Wegel, *Annals of the Amer. Rhin., Larg. and Ot. Soc.*, June, 1923, "Audiometric Methods and their Applications."
[58] Fletcher, "Nature of Speech and its Interpretation," *Jour. Frank. Inst.*, June, 1922.
[59] Lillie, R. S., "The Relation of Stimulation and Conduction in Irritable Tissues to Changes in Permeability of the Limiting Membranes," *Amer. J. Physiol.*, 28, 197–223, 1911. Also other papers.
[60] Lucas, K., "The Conduction of the Nervous Impulse," London, 1919.

Discussions of the Theory of Hearing

[61] Roaf, *Phil. Mag.*, 43, pp. 349–354, Feb., 1922.
[62] Wrightson, "Analytical Mechanism of Internal Ear," 1918.
[63] Morton, *Phys. Soc. Lond.*, 31, p. 101, Apr., 1919.
[64] Peterson, *Psych. Rev.*, 20, p. 312, 1913.
[65] Boring and Titchener, *Physiolog. Abs.*, 6, p. 27, April, 1921.
[66] Meyer, *Amer. Jour. of Psych.*, 18, pp. 170–6, 1907.
[67] Hartridge, *Brit. Jour. Psychol.*, 12, pp. 248–52, 1921.
[68] Hartridge, *Brit. Jour. Psychol.*, 12, pp. 277–88, 1921.
[69] Gray, *Jour. Anat. Phys.*, 34, p. 324, 1900.
[70] Hartridge, *Brit. Jour. Psychol.*, 12, pp. 142–6, 1921.
[71] Hartridge, *Brit. Jour. Psychol.*, 12, pp. 248–52, 1921.
[72] Hartridge, *Nature*, 107, pp. 394–5, May 26, 1921; 107, p. 204, April 14, 1921.
[73] Hartridge, *Brit. Jour. Psychol.*, 12, pp. 362–82, Aug. 16, 1921.

[74] Hartridge, *Nature*, 107, p. 811, Aug. 25, 1921.
[75] Hartridge, *Nature*, 109, p. 649, May 20, 1920.
[76] Ackerman, *Nature*, 109, p. 649, May 20, 1920.
[77] Hartridge, *Nature*, 110, pp. 9–10, July 1, 1922.
[78] Wilkinson, *Proc. of Roy. Soc. Med.*, 15, Sect. of Otol., pp. 51–3, 1922.
[79] Bayliss, *Nature*, 110, p. 632, Nov. 11, 1922.
[80] Broemser, *Physiol. Abs.*, 6, p. 28, April, 1921. (Abs. from *Sitzung. d. Ges. f. Morphol. u. Physiol. in Munchen*, p. 67, 1920.)
[81] Weiss, *Psychol. Rev.*, 25, p. 50, 1918.
[82] Abraham, *Ann. d. Phys.*, 60, No. 17, p. 55, 1919.
[83] Buck, *N. Y. Med. Jour.*, June, 1874.
[84] Bryant, Repr. Trans. of the *Amer. Otol. Soc. Transact.*, 1909.
[85] Marage, *Comp. Rend.*, 175, p. 724, Oct. 23, 1922.
[86] Rayleigh, *Sci. Abs.*, Sec. A., 22, p. 124, Mar. 31, 1919. (Abs. from *Nature*, 102, p. 304, Dec. 19, 1918.)
[87] Marage, *Comp. Rend.*, 172, p. 178, Jan. 17, 1921.
[88] Dahns, *Monasschr. f. Ohrenheilk. u. Laryng.-Rhinol.*, 56, p. 23, 1922.
[89] Barton, *Nature*, 110, pp. 316–9, Sept. 2, 1922.
[90] Scripture, *Nature*, 109, p. 518, Apr. 22, 1922.
[91] Ogden, *Psychol. Bull.*, 15, p. 76, 1918.
[92] Ogden, *Psychol. Bull.*, 16, p. 142, 1919.
[93] Ogden, *Psychol. Bull.*, 17, p. 228, 1920.

The Contributors to this Issue

George A. Campbell, B.S., Massachusetts Institute of Technology, 1891; A.B., Harvard, 1892; Ph.D., 1901; Göttingen, Vienna and Paris, 1893–96. Mechanical Department, American Bell Telephone Company, 1897; Engineering Department, American Telephone and Telegraph Company, 1903–1919; Department of Development and Research, 1919—; Research Engineer, 1908—. Dr. Campbell has published papers on loading and the theory of electric circuits and is also well-known to telephone engineers for his contributions to repeater and substation circuits. The electric filter which is one of his inventions plays a fundamental rôle in telephone repeater, carrier current and radio systems.

Robert W. King, A.B., Cornell University, 1912; Ph.D., 1915; assistant and instructor in physics, Cornell, 1913–17; Engineering Department of the Western Electric Company, 1917–20; Department of Development and Research, American Telephone and Telegraph Company, 1920–21; Information Department, 1921—. While with the Western Electric Company, Mr. King's work related to the design and construction of vacuum tubes and allied high vacuum apparatus.

Karl K. Darrow, S.B., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., in physics and mathematics, University of Chicago, 1917; Engineering Department, Western Electric Company, 1917—. At the Western Electric, Mr. Darrow has been engaged largely in preparing studies and analyses of published research in various fields of physics.

H. D. Arnold, Ph.B., Wesleyan, 1906; M.S., 1907; Ph.D., Chicago, 1911; assistant in physics, Wesleyan, 1906–07; Chicago, 1908; professor, Mt. Allison, 1909–10; Engineering Department of the Western Electric Company, Research Engineer, 1911—; Director of Research, 1923—. Dr. Arnold has been in direct charge of the development of the vacuum tube for telephone repeaters and radio purposes, and also other items of telephone equipment.

Lloyd Espenchied, Pratt Institute, 1909; United Wireless Telegraph Company as radio operator, summers, 1907–08; Telefunken Wireless Telegraph Company of America, assistant engineer, 1909–10;

American Telephone and Telegraph Company, Engineering Department and Department of Development and Research, 1910—. Took part in long distance radio telephone experiments from Washington to Hawaii and Paris, 1915; since then his work has been connected with the development of radio and carrier systems.

Harvey Fletcher, B.S., Brigham Young, 1907; Ph.D., Chicago, 1911; instructor of physics, Brigham Young, 1907–08; Chicago, 1909–10; Professor, Brigham Young, 1911–16; Engineering Department, Western Electric Company, 1916—. The present paper by Dr. Fletcher gives some of the results of an investigation which is being made of the relation between the frequency characteristics of telephone circuits and the intelligibility of transmitted speech. Dr. Fletcher has also published on Brownian movements, ionization and electronics.

VOLUME II OCTOBER, 1923 NUMBER 4

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per copy *$1.50 per year*

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the American Telephone and Telegraph Company, through its Information Department, in behalf of the Western Electric Company and the Associated Companies of the Bell System

Address all correspondence to the Editor
Information Department
AMERICAN TELEPHONE AND TELEGRAPH COMPANY
195 BROADWAY, NEW YORK, N. Y.

50c. Per Copy *$1.50 Per Year*

EXPLANATORY

Previous to 1907 there were maintained in the Bell System, three laboratories and departments of development, research and experiment,—one by the American Telephone and Telegraph Company at Boston, one by the Western Electric Company at Chicago and one by the Western Electric Company at New York.

In 1907, in the interest of economy and efficiency, these were consolidated so far as laboratory and experimental work were concerned and the Bell System Laboratory at Bethune and West Streets, New York, was established. The expense of operation is divided between the American Telephone and Telegraph Company and the Western Electric Company according to the nature of the work done.

In the Bell System the American Telephone and Telegraph Company undertakes, through constant association with the operating organizations, to formulate the requirements, present and future, of the Bell System. Out of these requirements come the problems of the American Telephone and Telegraph Company's Department of Development and Research and the System laboratory. After the problems have been satisfactorily solved, the Department of Development and Research adopts as standard the systems, equipment and apparatus thus produced which are then specified for their proper uses in the Associated Companies by the Engineering Department of the American Telephone and Telegraph Company. When different departments and companies are mentioned in this publication, so far as they are parts of the Bell System, they are parts of *one* working organization

H. B. THAYER, *President,*
American Telephone and Telegraph Company.